BMJ Open

Importance and added value of functional impairment in order to predict mortality: a prospective observation study in medical inpatients

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-014464
Article Type:	Research
Date Submitted by the Author:	26-Sep-2016
Complete List of Authors:	Torisson, Gustav; Lunds Universitet, Clinical Memory Research Unit Dept of Clinical Sciences Stavenow, Lars; Skanes universitetssjukhus Malmo, Department of Internal Medicine Minthon, Lennart; Clinical Memory Research Unit, Department of Clinical Sciences Londos, Elisabet; Clinical Memory Research Unit, Department of Clinical Sciences, Lund University
Primary Subject Heading:	Geriatric medicine
Secondary Subject Heading:	Research methods
Keywords:	GERIATRIC MEDICINE, STATISTICS & RESEARCH METHODS, Delirium & cognitive disorders < PSYCHIATRY

Importance and added value of functional impairment in order to predict mortality: a prospective observation study in medical inpatients

Authors

Gustav Torisson[1], PhD, Gustav.Torisson@med.lu.se
Lars Stavenow[2], associate professor, Lars.E.Stavenow@skane.se
Lennart Minthon[1], professor, Lennart.Minthon@skane.se
Elisabet Londos[1], associate professor, Elisabet.Londos@skane.se

Affiliations

[1] Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden.

[2] Department of Internal Medicine, Skåne University Hospital, Malmö Sweden.

Corresponding author:
Gustav Torisson, Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden.
Adress: Simrisbanvägen 14, 20502 Malmö.
E-mail: gustav.torisson@med.lu.se.
Telephone: +46 40 33 77 85

Word count: 2857

Abstract

Background: Accurate estimation of prognosis in multimorbid hospital patients could improve quality of care. This study aims to determine the relative importance and added value of a performance-based ADL (activities of daily living) measure with regard to mortality prediction.

Methods: Two hundred inpatients, aged over 60 years, were recruited at the Department of General Internal Medicine at a tertiary university hospital. Two nested survival models were built, one with established risk factors (age, sex, Charlson comorbidity index, haemoglobin, albumin, body mass index, and glomerular filtration rate)**,** and one using the same covariates with the GBS (Gottfries-Bråne-Steen) ADL measure added. The relative importance of GBS-ADL was evaluated in the full model. The added value of GBS-ADL was determined by comparing the nested models using four approaches: difference in overall χ^2, discrimination, continuous net reclassification index (NRI > 0) and integrated discrimination improvement (IDI).

Results: In the full model, GBS-ADL was the single most important predictor of mortality (χ^2 - df = 30, p <0.001). The likelihood ratio χ^2 test showed significant added value of ADL (p<0.001). The c statistic was 0.78 with ADL and 0.72 without, (difference 0.058, 95% CI = 0.022 to 0.094). The NRI > 0 was 0.42 (95% CI 0.20 to 0.58) and IDI 0.15 (95% CI 0.07 to 0.22).

Conclusions: Compared to a set of available clinical risk factors, impairment in ADL was a stronger predictor of all-cause mortality, showing substantial added value. Implementing quantitative ADL measurements could enable more appropriate and individual care for the elderly.

Keywords: aging, comorbidity, mortality, functional status, statistical modeling

Strengths and limitations of this study

- A rigorous statistical approach was used to determine the prognostic value of impaired ADL (activities of daily living) with regard to mortality in elderly inpatients.
- Impaired ADL, measured using a standardised performance-based scale, was shown to be a strong and independent marker of poor prognosis.
- However, the results need to be confirmed in other settings and for other ADL scales to be considered generalisable.

INTRODUCTION

Improving the accuracy of prognostic estimates could have several benefits for medical inpatients. Such benefits include reduced overtreatment, such as polypharmacy or the use of life-sustaining measures inconsistently with patients' preferences[1-4]. Other elderly patients are withheld treatment due to an incorrectly supposed poor prognosis, this could possibly be another important aspect[5-7]. Furthermore, patients with poor prognosis may prefer improved quality of life over extended survival. Therefore, accurate estimates could support doctors initiating a discussion regarding goals of care[8]. In addition, advance care planning could help patients and families to make necessary arrangements and increase quality of life[9-11].

Impairment in ADL (activities of daily living) is a well-known predictor of mortality and lower quality of life in hospitalized and community-dwelling elderly[12-20]. However, the majority of studies use interview-based scales[13, 15, 18], shown to differ significantly from performance-based ones[21, 22]. In addition, several studies use regression models without reporting overall performance[14, 15, 18, 23] and only few studies determine the added value of ADL[13, 15]. Recently, novel statistical methods have been introduced to establish the incremental value of prognostic markers[24].

In the present study, we aim to use these methods in order to determine the relative importance and added value of a performance-based ADL measure compared to clinical data, with regard to mortality prediction.

METHOD

This study constitutes a secondary analysis, all patients were concurrently taking part in a prospective trial, aiming to improve quality of care[25].

Setting

The study was carried out at the Department of General Internal Medicine at Skåne University Hospital in Malmö, Sweden. This teaching hospital provides care to the city's approximately 300,000 inhabitants. The department has four wards, with a total of 100 beds. Patients are admitted through the hospital's Emergency Department. Normally, the patients in the

department are elderly with multiple comorbidities. More specialized medical departments (cardiology, nephrology, endocrinology etc.) are separate and were not included in this study.

Patients

The recruitment of patients, that took place in 2009 and 2010, has been described in detail in a previous publication, including a flowchart[25]. In short, patients aged over 60 years, living in their own homes were eligible. Exclusion criteria comprised of terminal disease, language barrier, blindness/deafness/aphasia or other disease with inability to communicate, transfer to another department/ICU, early discharge and isolation due to communicable disease.

In total, 200 patients were included and underwent a baseline measurement. One half (101) of the patients constituted a control group while the other half (99) received a hospital-based, multidisciplinary intervention aiming to reduce rehospitalizations. The intervention consisted of a medication overview, improved discharge planning, telephone follow-up and improved liaison with GPs. Group allocation (intervention or control) used convenience sampling with geographic selection. At one-year follow-up, the intervention group had significantly fewer rehospitalizations than the control group[25].

ADL measurement

As part of the baseline measurement in the original trial, an ADL measurement was implemented by two experienced occupational therapists, who had received special training. The assessment was carried out when patients were stabilized, typically a few days into the admission.

The ADL subset of the GBS (Gottfries Bråne Steen) scale rates six items: dressing, food intake, physical activity, spontaneous activity, personal hygiene and toileting[26]. Items are scored on a performance-based 7-point scale ranging from 0 (best) to 6 (worst). For example, dressing is scored as follows:

0: Dresses and undresses without help

1:

2: Gets help with buttons, zips etc.

3:

4: Requires help from a caregiver to dress and undress but takes an active part

5:

6: Is completely dependent on a caregiver to be dressed and undressed

The points 1, 3 and 5 are not defined but are used by the observer to increase discrimination. Combining the six items gives a total ADL score of 0 (no impairment) to 36 (maximum impairment).

Other data from the original trial protocol

The Charlson comorbidity index was collected from the original protocol, to obtain a measure of combined comorbidity[27]. This index' performance concerning short-term and long-term mortality has recently been validated[28].

Data collection from medical records

Additional data was collected retrospectively regarding physiological and laboratory values. Since no blood samples were drawn in the original trial, only clinical data could be used. Candidate predictors were selected á priori on the basis of availability and previously established association with all-cause mortality. All data was obtained from the same hospital episode as ADL was measured. If unavailable during that hospitalization, the data point was labelled "missing". If several data points were found during the hospitalisation, the one closest to admission was used. The following variables, all independently related to all-cause mortality, were collected: Body mass index (BMI), kg/m^2, Hemoglobin (Hb), g/L, estimated Glomerular filtration rate (eGFR), ml/min, Albumin, g/L, Brain Natriuretic Peptide (BNP), ηg/L[29, 30, 31-33].

Statistical method

The goal of the present study was to compare the GBS-ADL measurement with the best set of available clinical risk factors using survival analysis. As a secondary analysis, no specific power calculation was done. First, we built a multivariate Cox regression model, called "model without ADL", using the established risk factors as covariates. Then, this model was refitted, with ADL added, to obtain the "full model". To determine the added value of ADL, the performance of these two models were compared. In addition, the relative importance of ADL was examined in the "full model".

The modelling algorithm is based on previous recommendations, primarily by Harrell et al and Steyerberg et al [24, 34-37]. All steps are explained in larger detail in Supplementary File 1. All modelling was executed in R, the script is supplied in Supplementary File 2.

1. **Outcome**. The study endpoint was mortality status on Feb 6th 2014. Follow-up was defined as time from discharge of the original hospitalisation.

2. **Crude analysis**. Separate bivariate proportional hazards regressions were carried out for all variables on their original scaling. Crude analysis were accomplished for all separate ADL items but in further analysis only the total GBS-ADL score was used.

3. **Missing data.** Missing values in covariates were quantified and controlled for systematic patterns resulting in their missing status. Missing values were then imputed using an imputational regression model.

4. **Variable transformations**: Haemoglobin was pre-specified to have a non-linear association with mortality. All other continuous variables were tested for non-linearity and transformed accordingly. Outliers were controlled for data entry errors and considered for truncation.

5. **Fitting the two multivariate models.** The "model without ADL" was fitted first, using the transformations and imputations described above. Then, ADL was added and the model was refitted to obtain the "full model".

6. **Multicolinearity.** The models were tested using the VIF (variance inflation factor).

7. **Interactions**: Pooled two-way interaction tests were carried out for all variables, in both models, separately. If the pooled test was significant, specific interactions were pursued for that variable.

8. **Proportional hazards**. The proportional hazards assumption was tested with global tests and Schoenfeld residual plots for each variable.

9. **Influential observations.** Observations with a standardised DfBeta > 0.20 standard errors were noted for each variable. As ADL was of particular interest, a sensitivity analysis was performed without this variable's influential observations.

10. **Determining the relative importance of ADL.** An ANOVA test was performed to determine the relative contribution of the separate variables, including their interaction terms and non-linear terms.

11. **Determining added value of ADL**. To determine added value, the "model without ADL" and the "full model" were compared using:

    a. **Likelihood ratio test.** Performed as a χ^2 testing the difference in Likelihood ratio between the models' χ^2 over df = number of additional independent variables.

    b. **Discrimination**, measured with the C, or concordance, statistic. The C statistic is the probability that, in a case-control pair, the case will be given a higher predicted risk from the model than the control. C statistics ranges from 0.5 (coin toss, useless) to 1.0 (perfect discrimination). The difference in C statistic between models was tested using the method described by Uno et al.[38] .

    c. **NRI >0 (Continuous net reclassification index)**[39, 40]. This index determines to what extent adding a new variable leads to a change in the correct direction of predicted risk for each observation (towards higher risk for deceased, towards lower for survivors). NRI ranges from 0 (no increased value, useless) to 1(all cases reclassified in the right direction). NRI has been shown to be more sensitive than change in C index, especially when the baseline model has a good performance.

    d. **IDI**. Integrated discrimination improvement. Originally developed by Pencina et al. for logistic models, IDI has been extended to time-to-event data[39, 41]. While NRI>0 measures the percentage of observations that have been reclassified, it cannot distinguish between a small change in prediction and a large. IDI, however, measures the mean amount of such change. IDI and NRI with confidence intervals were calculated with the method by Uno et al.[42]

12. **Internal validation**. Both models were internally validated through 1000 bootstrap resamples to estimate the amount of overfitting and to obtain optimism-corrected performance estimates.

13. **Updating and presenting final model**. The "full model" was updated through the use of a LASSO (least absolute shrinkage and selection operator) procedure to reduce the effects of overfitting[43, 44]. The updated LASSO model was used to build a nomogram, with which patients were stratified into four equally sized risk groups, displayed in a Kaplan-Meier graph.

RESULTS

In two cases, mortality status could not be obtained; these were discarded from further analysis. Of the remaining 198 cases, 126 were deceased at follow-up. The median follow-up time for survivors was 1428 days (range 1312-1548). Baseline characteristics are displayed in table 1.

Table 1. Baseline characteristics

Continuous variables	**mean (SD)**	**median (IQR)**	**min-max**
Age	83.4 (8.1)	85 (78-89)	60-100
Charlson comorbidity index	2.3 (1.5)	2 (1-3)	0-7
GBS-ADL, total	6.8 (5.7)	5 (2-10)	0-25
GBS-ADL, dressing	1.3 (1.4)	1(0-2)	0-5
GBS-ADL, food intake	0.1 (0.4)	0(0-0)	0-2
GBS-ADL, physical activity	2.0 (1.1)	2(2-2)	0-5
GBS-ADL, spontaneous activity	1.0 (1.2)	1(0-2)	0-5
GBS-ADL, hygiene	1.4 (1.4)	2(0-2)	0-5
GBS.ADL, toilet	0.9 (1.4)	0(0-1)	0-6
Hemoglobin, g/L	123 (19)	124 (111-136)	53-179
eGFR, ml/min, n = 197	42.3 (25)	37(26-51)	6-198
BMI, kg/m2, n = 195	24.7 (5.1)	24 (21-27)	14-42
Albumin, g/L, n = 181	31.5 (4.9)	32 (29-35)	14-42
BNP, ηg/L, n = 85	261 (297)	147 (54-377)	3-1618
Categorical variables	**number**	**percentage**	
Male sex	70	35%	
In intervention group in original study	99	50%	

Table 1. Baseline characteristics for the entire sample. n = 200 unless otherwise stated. ADL = activities of daily living, eGFR =estimated glomerular filtration rate, BMI = Body mass index, BNP = Brain natriuretic peptide.

The results from the crude analysis are presented in table 2.

Table 2. Crude analysis

Predictor	β	S.E	Wald X^2	p value	HR (95% CI)
GBS-ADL-total, points	0.08	0.013	37.8	<0.001	1.08 (1.06 - 1.11)
GBS-ADL-hygiene, points	0.38	0.06	37.7	<0.001	1.46 (1.29 - 1.65)
GBS-ADL-physical, points	0.46	0.08	36.0	<0.001	1.59 (1.36 - 1.84)
GBS-ADL-dressing, points	0.31	0.06	30.0	<0.001	1.36 (1.22 - 1.52)
eGFR, ml/min, n = 197	-0.029	0.005	29.3	<0.001	0.97 (0.96 - 0.98)
GBS-ADL-spontaneous, points	0.33	0.06	27.0	<0.001	1.40 (1.23 - 1.58)
Charlson index, points	0.22	0.06	15.2	<0.001	1.25 (1.18 - 1.40)
Hemoglobin, g/L	-0.019	0.005	14.6	<0.001	0.98 (0.97 - 0.99)
Albumin, g/L, n = 181	-0.064	0.018	13.1	<0.001	0.94 (0.90 - 0.97)
GBS-ADL- toileting, points	0.19	0.05	11.6	<0.001	1.20 (1.08 - 1.34)
Age, years	0.036	0.011	10.1	0.001	1.04 (1.01 - 1.06)
BMI, kg/m^2 , n = 195	-0.053	0.020	7.4	0.007	0.95 (0.91 - 0.99)
BNP, ηg/L, n = 85	0.0009	0.0003	6.7	0.01	1.001 (1 - 1.002)
ADL - food intake, points	0.34	0.21	2.7	0.10	1.41 (0.93 - 2.12)
Sex (0 = Female, 1 = Male)	0.29	0.18	2.6	0.11	1.34 (0.94 - 1.92)
Group in original study (0=control, 1=intervention)	0.11	0.18	0.37	0.54	1.12 (0.78 - 1.59)

Table 2. Crude Cox regression for all predictors, sorted by decreasing strength of association. S.E = standard error, HR = Hazard ratio, ADL = Activities of daily living, eGFR = glomerular filtration rate, BMI = body mass index, BNP = brain natriuretic peptide.

BNP was missing in 115 cases (58%) and the variable was discarded from further analysis. eGFR and BMI were missing in 1 and 3 cases, respectively; these were considered to be missing completely at random. Albumin was missing in 17 cases, these were predominantly female (15/17) and had lower scores on Charlson comorbidity index. Missing values were imputed with a minimal change in variable properties, see Supplementary File 1.

Hemoglobin was fitted using a 4-knot restricted spline and GBS-ADL was transformed using the natural logarithm. No other predictors showed significant non-linear properties and they were kept in their original form. eGFR had one extreme outlier at 198ml/min that was winsorized at the 99th percentile (118 ml//min).

A significant sex * BMI interaction was found and included into the models (low BMI was a significant predictor in men but not in women). Another interaction, eGFR * ADL, was included as well (ADL was a stronger predictor when eGFR was unimpaired and vice versa). No other significant interactions were found. No significant multicolinearities were found. The proportional hazards assumption was not violated. In the full model, 21 observations were influential, of which 9 for ADL and/or its interaction with eGFR. A sensitivity analysis with these cases removed showed a slight improvement in model fit and is presented in Supplementary File 1. However, all observations were kept in the models.

In the "full model", ADL was by far the most significant predictor. The relative importance of the predictor variables are shown in figure 1. All four measurements showed added value for model with ADL, see table 3.

Table 3. Added value of ADL

Model comparison	model without ADL	model with ADL	p value
Nagelkerke R^2	0.33	0.46	
Likelihood ratio χ^2	78.4 (11df)	121.0(13df)	<0.001
c statistic (95% CI)	0.72(0.67-0.76)	0.78(0.73-0.82)	0.001
½ NRI > 0 (95% CI)		0.42(0.20-0.58)	<0.001
IDI (95% CI)		0.15(0.07-0.22)	<0.001

Table 3. Comparison of the two nested survival models. NRI > 0 = continuous Net Reclassification Index, IDI = Integrated Discriminatory Improvement.

When bootstrapped 1000 times, the calibration slope of the "model without ADL" was 0.84 and of the "full model" 0.83. Optimism-corrected R^2 was 0.27 vs. 0.40, respectively. Optimism-corrected c statistics were 0.70 and 0.76. When the LASSO was employed to shrink coefficients and update the model, the mean shrinkage was 0.84. The nomogram using the updated model coefficients is shown in Supplementary File 1 and the subsequent Kaplan-Meier graph for the four risk groups are presented in figure 2.

DISCUSSION

In this study, we confirm that impaired ADL is an important predictor of mortality in elderly medical inpatients. The relative contribution of ADL was larger than of the available predictors in a real-life setting, including a comorbidity index, available physiological parameters and laboratory values. In addition, ADL showed a substantial added value when compared to a model combining all of these traditional predictors.

In the crude analysis, four of the GBS-ADL items were stronger predictors than the Charlson comorbidity index. Thus, a simple rating of dressing ability had better predictive value than a combined comorbidity measure, designed to predict mortality. This indicates that performance-based ADL measures are truly important mortality predictors in multimorbid patients. In multivariate analysis, impairment in ADL was by far the most important predictor and all four measures signaled added value when GBS-ADL was added to the traditional predictors.

The mechanism underlying the association between ADL and mortality is probably multifactorial. Impairment in ADL could contribute directly to mortality in some aspects. Obvious complications to functional decline include pressure sores, atrophy, falls, thrombosis etc. However, less intuitive factors could also apply, such as attaining multi-resistant bacteria or Clostridium Difficile[45 46]. Even more likely, ADL acts a proxy for a confounder not measured by the model. A possible such confounder is frailty, defined as an increased vulnerability, where small stressors lead to adverse outcomes, such as hospitalization or death[47]. The frailty phenotype includes unintentional weight loss, along with loss of strength, low physical activity, slow walking speed and exhaustion[48]. There is a considerable overlap between frailty, comorbidity and ADL impairment. Our study utilized specific measures for comorbidity and ADL impairment, but not for frailty. However, our model is most likely describing the effects of frailty as well.

Several methodological issues need to be addressed. First, the choice of ADL scale, where the GBS scale was chosen in order to facilitate implementation locally. There are large variations and lack of standardization regarding functional measures used in medical inpatients[49]. The GBS scale proved feasible and has been shown to have a good construct validity and inter-rater reliability[50]. In addition, the GBS-ADL has correlated strongly with other ADL measurements, for example Katz' index[51, 52]. A potentially confounding issue was the

concurrent non-randomized trial. However, the variable "control/intervention status" was included in all analyses, without any sign of bias. In addition, the sample size was small, and internal validation showed that our models were indeed overfitted, with a calibration slope of 0.83. This overfitting is probably not a result of having too many covariates but rather a result of the global interaction tests and tests of non-linearity. This multiple comparison situation has been called "testimation bias"[37]. The overall aim was not to develop the most comprehensive and parsimonious prediction model to use in future populations but to describe the importance and added value of ADL. Therefore, we prioritized not to miss clinically important interactions and/or transformations in the trade-off with overfitting. To compensate partly, we used a LASSO procedure to shrink estimates. The small sample size and the aim to compare ADL with the best possible model was also the reason underlying the imputation of missing values.

The primary strength of this study is the rigorous statistical approach. State-of-the-art methods were used in the model building to handle missing data, to address non-linearity, to screen for interactions, for model diagnostics and for internal validation. In addition, four different methods were applied to estimate added value. Previously, a study has showed increase in model χ^2 when adding a composite ADL measure, regarding 2-year mortality [15]. However, this study compared ADL only with comorbidity indices. With such a limited reference model it is likely that a new measure will add value but the final model could still perform poorly, which was reflected by low model χ^2 values and a final c statistic of 0.66. The use of comorbidity indices only as reference model is also far from the clinical reality. Another study shows increase in discrimination when adding an ADL measurement to a 1-year logistic regression mortality prediction model[13]. This study also starts with comorbidity indices alone and does not report any other measurement of overall performance (such as overall χ^2 or R^2). Our study compares ADL to a much more complex reference model and yet shows added value using both these previously applied measurements as well as several others.

Implications for further research include research regarding performance-based ADL scales, including the relation to specific frailty ratings. Larger studies could obtain head-to-head comparisons of ADL vs. disease-specific predictors, such as ejection fraction in heart failure.

Today, ADL is very often assessed in a variety of ways in medical inpatients, to assess the individuals' needs after discharge. Implementing a performance-based quantitative

measurement could have many benefits, also apart from prognostic value, such as increased standardization and the possibility to follow a patient over time. As a final remark, mortality prediction is not all about avoiding overtreatment due to a poor prognosis. Our model identified 50 elderly multimorbid medical inpatients with a 90% chance of 3-year survival. This group should not be undertreated simply due to age discrimination.

In conclusion, an ADL measurement showed significant added value as a predictor of mortality in a multimorbid elderly hospital population. Implementation of standardized ADL measurements could lead to better prognostic estimates and in the end a more appropriate and individualised care for the elderly.

DECLARATIONS

List of abbreviations

ADL: activities of daily living

GBS: Gottfries-Bråne-Steen

BMI: body mass index

eGFR: estimated glomerular filtration rate

BNP: brain natriuretic peptide

NRI: net reclassification index

IDI: integrated discrimination improvement

Ethics approval

All patients enrolled in the original study gave written informed consent. Both the original trial and the secondary analysis have been approved by the regional ethics committee at Lund University.

Availability of data and materials

Since the participatns were not specifically asked for consent to share data, such sharing is not compatible with the current Swedish legislation. The data protection officer at Skåne University hospital, the data protection officer at Lund university as well as lawyers at the Swedish data protection authority have unanimously adviced us not to publish data, even if anonymized.

Conflicts of interests

The authors declare no financial relationships with any organizations that might have an interest in the submitted work and no other relationships or activities that could appear to have influenced the submitted work.

Funders

This study was financed by the Swedish Research Council (Vetenskapsrådet #523-2010-520), the Swedish Brain Power programme, the National Swedish Board of Health and Welfare and the Governmental Funding of Clinical Research within the National Health Services (ALF). All researchers acted independently to the funding bodies. The funding agencies had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Authors' contributions

The study was conceived by LS and LM. Data collection was done by GT, supervised by EL. GT performed all analysis and drafted the manuscript, which was critically revised by LS, LM and EL. All authors have approved the final version.

Acknowledgments

The authors would like to thand Sofia Raccuia, Jenny Cappelin and Anna Johansson for their assistance in the data acquisition.

REFERENCES

1. Frankl D, Oye RK, Bellamy PE. Attitudes of hospitalized patients toward life support: a survey of 200 medical inpatients. The American journal of medicine 1989;**86**(6):645-8
2. Heyland DK, Barwich D, Pichora D, et al. Failure to engage hospitalized elderly patients and their families in advance care planning. JAMA internal medicine 2013;**173**(9):778-87.
3. Somogyi-Zalud E, Zhong Z, Hamel MB, et al. The use of life-sustaining treatments in hospitalized persons aged 80 and older. Journal of the American Geriatrics Society 2002;**50**(5):930-4
4. Teno JM, Gozalo PL, Bynum JP, et al. Change in end-of-life care for Medicare beneficiaries: site of death, place of care, and health care transitions in 2000, 2005, and 2009. JAMA : the journal of the American Medical Association 2013;**309**(5):470-7.
5. Biganzoli L, Wildiers H, Oakman C, et al. Management of elderly patients with breast cancer: updated recommendations of the International Society of Geriatric Oncology (SIOG) and European Society of Breast Cancer Specialists (EUSOMA). The Lancet Oncology 2012;**13**(4):e148-60.
6. Tahir M, Robinson T, Stotter A. How not to neglect the care of elderly breast cancer patients? Breast 2011;**20**(4):293-6.
7. Wildiers H, Kunkler I, Biganzoli L, et al. Management of breast cancer in elderly individuals: recommendations of the International Society of Geriatric Oncology. The Lancet Oncology 2007;**8**(12):1101-15.
8. Dev S, Clare RM, Felker GM, et al. Link between decisions regarding resuscitation and preferences for quality over length of life with heart failure. European journal of heart failure 2012;**14**(1):45-53.
9. Steinhauser KE, Christakis NA, Clipp EC, et al. Factors considered important at the end of life by patients, family, physicians, and other care providers. JAMA : the journal of the American Medical Association 2000;**284**(19):2476-82
10. Detering KM, Hancock AD, Reade MC, et al. The impact of advance care planning on end of life care in elderly patients: randomised controlled trial. BMJ 2010;**340**:c1345.
11. Wright AA, Zhang B, Ray A, et al. Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. JAMA : the journal of the American Medical Association 2008;**300**(14):1665-73.
12. Han PK, Lee M, Reeve BB, et al. Development of a prognostic model for six-month mortality in older adults with declining health. Journal of pain and symptom management 2012;**43**(3):527-39.
13. Covinsky KE, Justice AC, Rosenthal GE, et al. Measuring prognosis and case mix in hospitalized elders. The importance of functional status. Journal of general internal medicine 1997;**12**(4):203-8
14. Fried LP, Kronmal RA, Newman AB, et al. Risk factors for 5-year mortality in older adults: the Cardiovascular Health Study. JAMA : the journal of the American Medical Association 1998;**279**(8):585-92
15. Inouye SK, Peduzzi PN, Robison JT, et al. Importance of functional measures in predicting mortality among older hospitalized patients. JAMA : the journal of the American Medical Association 1998;**279**(15):1187-93
16. Lee SJ, Lindquist K, Segal MR, et al. Development and validation of a prognostic index for 4-year mortality in older adults. JAMA : the journal of the American Medical Association 2006;**295**(7):801-8.

17. Teno JM, Harrell FE, Jr., Knaus W, et al. Prediction of survival for older hospitalized patients: the HELP survival model. Hospitalized Elderly Longitudinal Project. Journal of the American Geriatrics Society 2000;**48**(5 Suppl):S16-24
18. Walter LC, Brand RJ, Counsell SR, et al. Development and validation of a prognostic index for 1-year mortality in older adults after hospitalization. JAMA : the journal of the American Medical Association 2001;**285**(23):2987-94
19. Torisson G, Minthon L, Stavenow L, et al. Cognitive impairment is undetected in medical inpatients: a study of mortality and recognition amongst healthcare professionals. BMC geriatrics 2012;**12**:47.
20. Torisson G, Stavenow L, Minthon L, et al. Reliability, validity and clinical correlates of the Quality of Life in Alzheimer's disease (QoL-AD) scale in medical inpatients. Health and quality of life outcomes 2016;**14**:90.
21. Magaziner J, Zimmerman SI, Gruber-Baldini AL, et al. Proxy reporting in five areas of functional status. Comparison with self-reports and observations of performance. American journal of epidemiology 1997;**146**(5):418-28
22. Sherman SE, Reuben D. Measures of functional status in community-dwelling elders. Journal of general internal medicine 1998;**13**(12):817-23
23. Reuben DB, Rubenstein LV, Hirsch SH, et al. Value of functional status as a predictor of mortality: results of a prospective study. The American journal of medicine 1992;**93**(6):663-9
24. Steyerberg EW, Pencina MJ, Lingsma HF, et al. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. European journal of clinical investigation 2012;**42**(2):216-28.
25. Torisson G, Minthon L, Stavenow L, et al. Multidisciplinary intervention reducing readmissions in medical inpatients: a prospective, non-randomized study. Clinical interventions in aging 2013;**8**:1295-304.
26. Gottfries CG, Brane G, Gullberg B, et al. A new rating scale for dementia syndromes. Archives of gerontology and geriatrics 1982;**1**(4):311-30
27. Charlson M, Szatrowski TP, Peterson J, et al. Validation of a combined comorbidity index. Journal of clinical epidemiology 1994;**47**(11):1245-51
28. Frenkel WJ, Jongerius EJ, Mandjes-van Uitert MJ, et al. Validation of the Charlson Comorbidity Index in acutely hospitalized elderly adults: a prospective cohort study. Journal of the American Geriatrics Society 2014;**62**(2):342-6.
29. Janssen I, Katzmarzyk PT, Ross R. Body mass index is inversely related to mortality in older people after adjustment for waist circumference. Journal of the American Geriatrics Society 2005;**53**(12):2112-8.
30. Culleton BF, Manns BJ, Zhang J, et al. Impact of anemia on hospitalization and mortality in older adults. Blood 2006;**107**(10):3841-6.
31. Tonelli M, Wiebe N, Culleton B, et al. Chronic kidney disease and mortality risk: a systematic review. Journal of the American Society of Nephrology : JASN 2006;**17**(7):2034-47.
32. Goldwasser P, Feldman J. Association of serum albumin and mortality risk. Journal of clinical epidemiology 1997;**50**(6):693-703
33. Wallen T, Landahl S, Hedner T, et al. Brain natriuretic peptide predicts mortality in the elderly. Heart 1997;**77**(3):264-7
34. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine 1996;**15**(4):361-87.

35. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. European heart journal 2014;**35**(29):1925-31.
36. Harrell Jr FE. *Regression Modeling Strategies - With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer-Verlag, 2001.
37. Steyerberg EW. *Clinical Prediction Models - A Practical Approach to Development, Validation, and Updating*. New York: Springer, 2009.
38. Uno H, Cai T, Pencina MJ, et al. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Statistics in medicine 2011;**30**(10):1105-17.
39. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Statistics in medicine 2008;**27**(2):157-72; discussion 207-12.
40. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Statistics in medicine 2011;**30**(1):11-21.
41. Chambless LE, Cummiskey CP, Cui G. Several methods to assess improvement in risk prediction models: extension to survival analysis. Statistics in medicine 2011;**30**(1):22-38.
42. Uno H, Tian L, Cai T, et al. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. Statistics in medicine 2013;**32**(14):2430-42.
43. Tibshirani R. Regression shrinkage and selection via the Lasso. J Royal Statist Soc B 1996;**58**(1):267-88
44. Tibshirani R. The lasso method for variable selection in the Cox model. Statistics in medicine 1997;**16**(4):385-95
45. Chen TY, Anderson DJ, Chopra T, et al. Poor functional status is an independent predictor of surgical site infections due to methicillin-resistant Staphylococcus aureus in older adults. Journal of the American Geriatrics Society 2010;**58**(3):527-32.
46. Rao K, Micic D, Chenoweth E, et al. Poor functional status as a risk factor for severe Clostridium difficile infection in hospitalized older adults. Journal of the American Geriatrics Society 2013;**61**(10):1738-42.
47. Clegg A, Young J, Iliffe S, et al. Frailty in elderly people. Lancet 2013;**381**(9868):752-62.
48. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. The journals of gerontology Series A, Biological sciences and medical sciences 2001;**56**(3):M146-56
49. Buurman BM, van Munster BC, Korevaar JC, et al. Variability in measuring (instrumental) activities of daily living functioning and functional decline in hospitalized older medical patients: a systematic review. Journal of clinical epidemiology 2011;**64**(6):619-27.
50. Brane G, Gottfries CG, Winblad B. The Gottfries-Brane-Steen scale: validity, reliability and application in anti-dementia drug trials. Dementia and geriatric cognitive disorders 2001;**12**(1):1-14
51. Dencker K, Gottfries CG. Activities of daily living ratings of elderly people using Katz' ADL Index and the GBS-M scale. Scandinavian journal of caring sciences 1995;**9**(1):35-40
52. Katz S, Ford AB, Moskowitz RW, et al. Studies of Illness in the Aged. The Index of Adl: A Standardized Measure of Biological and Psychosocial Function. JAMA : the journal of the American Medical Association 1963;**185**:914-9

Figure Legends

Figure 1. Relative importance of predictors in the multivariate “model without ADL” and the “full model”. Interaction terms and non-linear effects have been incorporated in the variables. control = the grouping variable from the original study. BMI = body mass index, eGFR = estimated glomerular filtration rate.

Figure 2. Kaplan-Meier estimates from the updated full model including customary risk factors and ADL. The participants have been stratified into four equally sized groups by quartiles of risk.

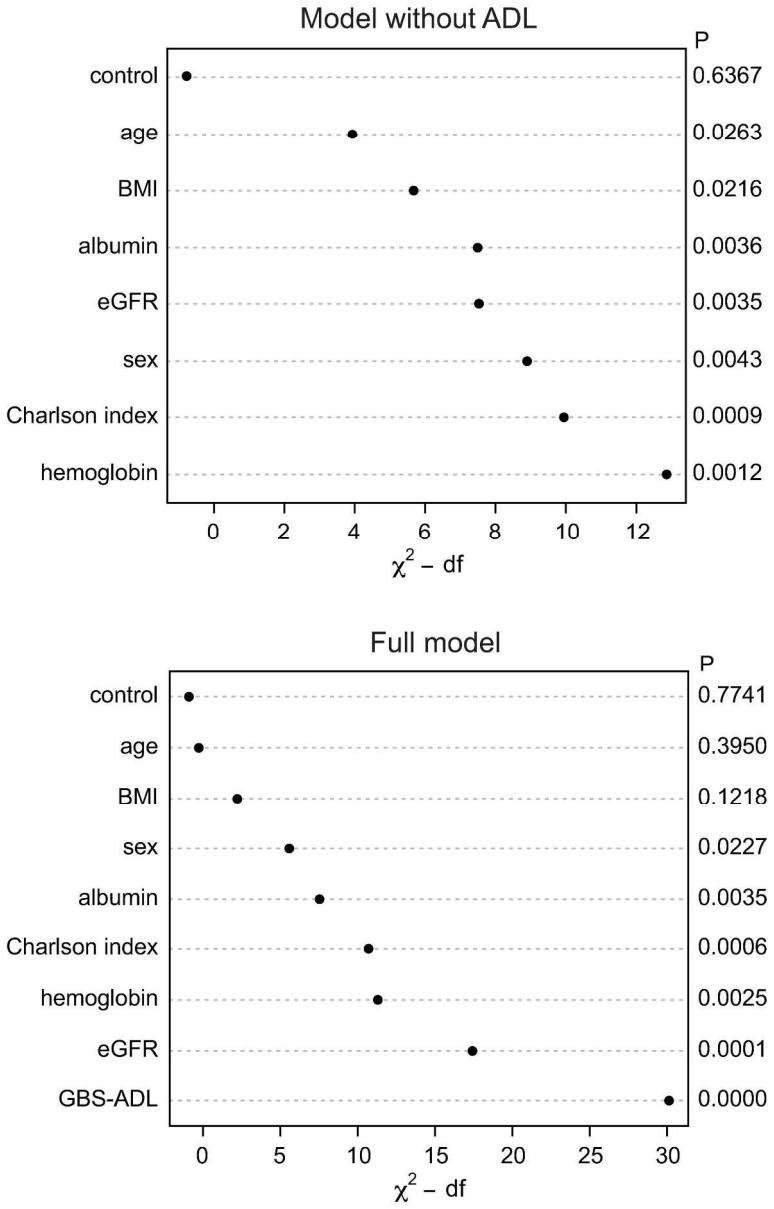

Figure 1. Relative importance of predictors in the multivariate "model without ADL" and the "full model". Interaction terms and non-linear effects have been incorporated in the variables. A higher χ2-df value indicates a stronger association. Control = the grouping variable from the original study. BMI = body mass index, eGFR = estimated glomerular filtration rate.

189x298mm (300 x 300 DPI)

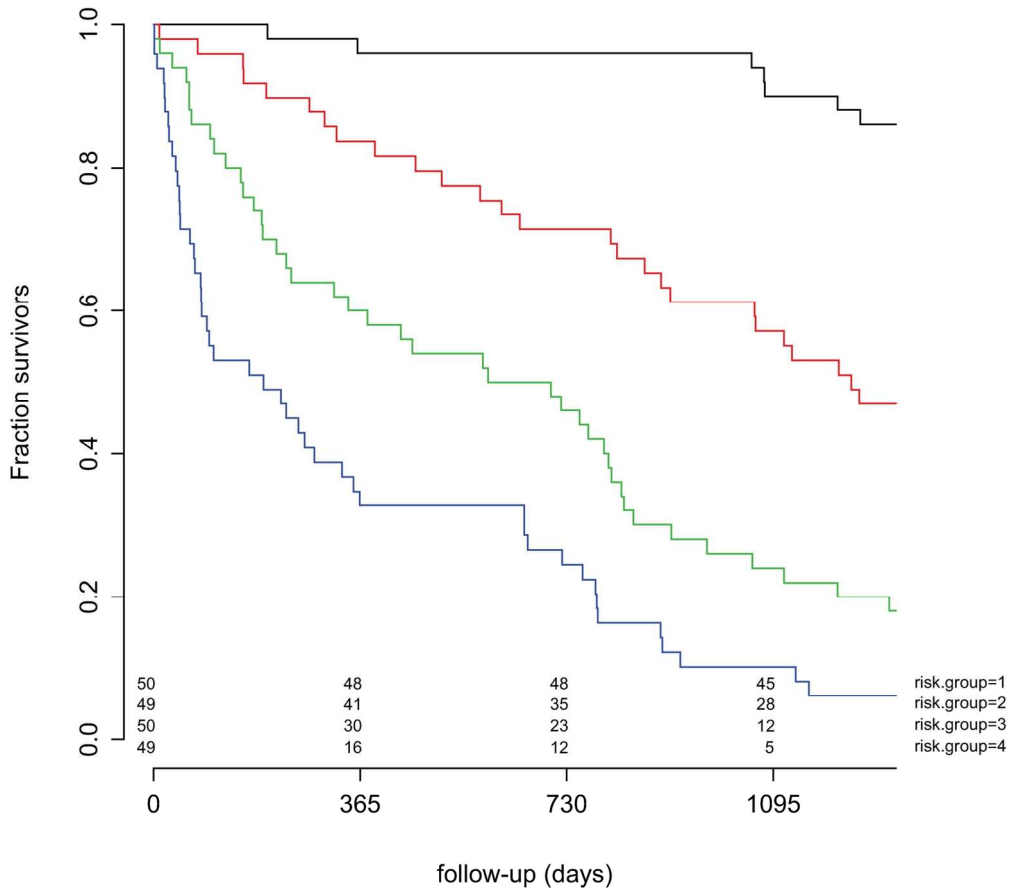

Figure 2. Kaplan-Meier estimates from the updated full model including customary risk factors and ADL. The participants have been stratified into four equally sized groups by quartiles of risk.

149x132mm (300 x 300 DPI)

Appendix 1 –statistics

General aspects

The overall aim was to compare ADL with the best possible model containing customary risk factors.
To achieve this, two models were developed and compared, the "model without ADL" and the "full model".
Data was originally stored in an SPSS file. All data analysis was performed in R [1]. Code is provided in appendix 2.

1. Outcome

Survival status was determined using the local region's electronic registry on the 6th February 2014. The time variable was defined as days from discharge to death or censoring at study endpoint, whichever came first. Those surviving at endpoint had been followed for a median of 1428 days (range 1312-1548). The baseline survival function is shown in figure e1.

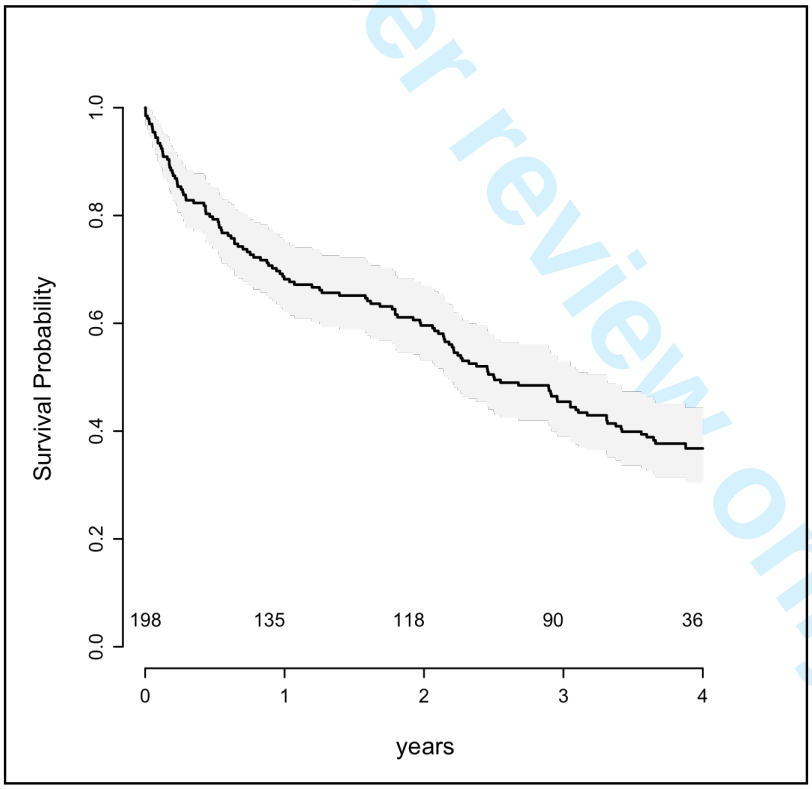

figure e1. Baseline survival function.

Generally, it is important to describe the quantity of cases with missing outcome and to determine if there are any underlying patterns. Otherwise, simple exclusion may affect representativity [2].

In our study, two cases were missing survival status due to having moved abroad (no longer in the region's registry) Hypothetically, these cases could be assumed to be in better health

(severely diseased patients are unlikely to move abroad?). However, they were considered too few to affect representativity and were discarded from further analysis. Thus, the number of cases decreased from 200 to 198.

2. Crude analysis

To obtain a first estimate of the effect of the variables/predictors , a crude analysis was performed. Bivariate Cox proportional hazards regressions were carried out separately for all variables, including only outcome and the variable. All variables were treated in their original form, on their original scale. Observations with missing values were excluded from crude analysis. Data is presented with β coefficients, Standard errors, Wald χ^2, p value and hazard ratios in table 2 in the article.

In the crude analysis, all variables/potential predictors were statistically significant except sex, control/intervention status in the original study and the ADL item "food intake". Regarding the latter, the distribution was severely skewed, with only 18 cases (9%) having a non-zero value. To obtain a preliminary ranking of importance, the variables were sorted by decreasing Wald χ^2 in the table in the article.

In crude analysis, all separate GBS-ADL items were included but in further multivariate analysis, only the total GBS-ADL score was used, to avoid fitting too many variables and multicolinearity (the ADL items were intercorrelated at r = 0.8-0.9)

3. Missing data

In general, it is important to analyse missing data patterns in predictors. The first step is to determine the quantity of missing data. The second step is whether data is missing completely at random or if there are underlying patterns. When these prerequisites have been fulfilled, there are several approaches to missing data:

1. Listwise deletion, discarding all observations with any missing data points. The advantage of this approach is that no "manipulation" is done. Therefore, this method may seem intuitively most correct. The obvious disadvantage is that sample size could be substantially diminished. In addition, representativity could be affected, if missing a variable is systematically associated with other characteristics.

2. Using simple imputation. This technique substitutes missing values with the mean, mode or median value. This could be acceptable only if the variable is missing completely at random and the percentage of missing values small.

3. Using a more complex imputational technique. This approach uses customised regression models including all other covariates to obtain a stable prediction of the missing values. This method has been described and emphasized in several publications [3-7].

When analysing the quantity of missing data, eGFR was missing in one case, BMI in three cases. Albumin was missing in 17 (9%) cases. BNP was missing in 113 (56%) observations.

The BNP variable was discarded from further analysis, as it had more than 50% missing. BMI and eGFR were considered missing completely at random. However, we found that cases

with missing albumin were predominantly female (15 female vs. 2 male, $\chi^2 = 3.32$, p = 0.056) and had lower score on Charlson comorbidity index (1.47 vs. 2.33, F = 11.3, p = 0.002). Thus, excluding cases with missing albumin would affect representativity. Discarding the albumin variable would affect the overall aim, to compare ADL with the best possible traditional model. Therefore, the missing values in BMI, eGFR and Albumin were imputed using a single conditional imputation method (with the `transcan` function in R). In total, the effect of imputations was very small on the variable properties, as shown below.

Variable	β	S.E	Wald X2	p value	HR (95% CI)
Albumin, g/L, n = 181	-0.064	0.018	13.1	<0.001	0.94 (0.90 - 0.97)
- with imputation, transcan	-0.066	0.017	14.7	<0.001	0.94 (0.91 - 0.97)
eGFR, ml/min, n = 197	-0.029	0.005	29.3	<0.001	0.97 (0.96 - 0.98)
- with imputation, transcan	-0.029	0.005	29.2	<0.001	0.97 (0.96 - 0.98)
BMI, kg/m^2 , n = 195	-0.053	0.020	7.4	0.007	0.95 (0.91 - 0.99)
- with imputation, transcan	-0.053	0.020	7.4	0.006	0.95 (0.91 - 0.99)

Table e1. Effect of imputation on variable properties.

4. Variable considerations

Extreme outliers

In regression, outliers may be defined as observations with more than 3 interquartile ranges over the third quartile or below the 1st quartile. Such extreme values may affect a regression model significantly. First data entry errors should be considered and pursued. Then the biological plausibility should be considered. If plausible we may consider a truncation at the 99th or 1st percentile [8].

In our study, data screening revealed, that for eGFR there was one extreme outlier with an estimated value of 198 ml /min (> 6 IQR over 3rd quartile), see boxplot.

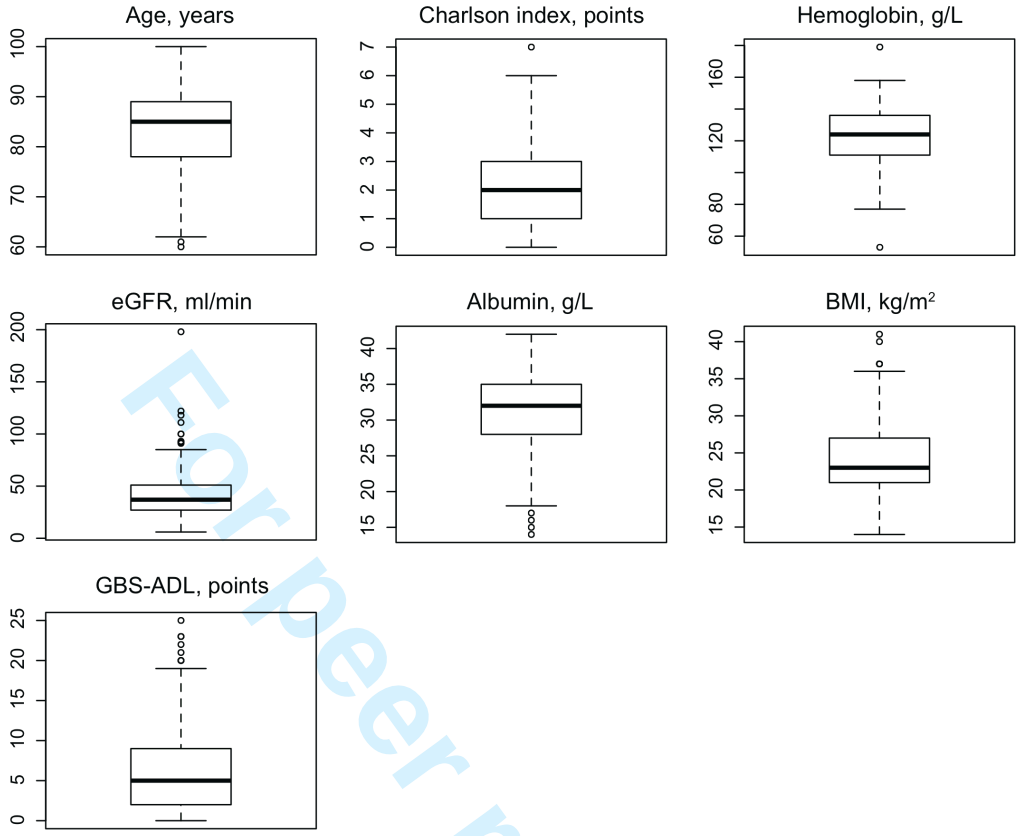

Figure e2. Boxplots of the continuous predictors. eGFR = Glomerular filtration rate, BMI = Body mass index, ADL = Activities of daily living.

This case was screened for data entry errors but none were found. Regarding biological plausibility, eGFR was measured with the Cockcroft-Gault formula ((140-age) * weight * constant)/Serum Creatinine in μmol/L, where the constant is 1.23 for men and 1.04 for women. Thus, GFR was not measured directly, but estimated and sensitive to extreme values in both serum creatinine, age and body weight. With this reservation, we considered the value to be biologically plausible. However, we did not consider it clinically important to compare one elderly patient with 198 ml/min in eGFR with another with 120 ml/min with regard to mortality. Therefore, eGFR was winsorized at the 99th percentile (118 ml/min). This led to a slightly improved fit in univariate performance.

Variable	β	S.E	Wald X2	p value	HR (95% CI)
eGFR, ml/min	-0.029	0.005	29.2	<0.001	0.97 (0.96 - 0.98)
- winsorized at 99th percentile	-0.029	0.005	29.8	<0.001	0.97 (0.96 - 0.98)

Table e2. effect of winsorization on variable properties.

Non-linearity

Most regression model assume that the predictors are linearly related to the outcome. However, non-linear relationships, such as U-shapes, for continuous variables are common.

There are several ways to address non-linearity:

First, assuming that the variable is linear. The advantage of this approach is that it results in an easily interpreted main effect, for example the Hazard Ratio in survival analysis. This is

the approach used in our crude comparisons. However, the approach is potentially problematic. For hemoglobin, this would mean that the risk difference between two individuals with 170 and 130 g/L would be the same as between two with 90 and 50, respectively. In addition, this approach cannot handle U-shaped risks, it is likely that someone with 200 g/L in Haemoglobin with dehydration or polycytemia does not have better survival than someone with 140 g/L

Second, to dichotomize the variable, using a previously established cut-off, is another frequently used approach. However it is not recommended as it ignores a lot of information [9]. In our example, applying the WHO cut-off for anemia (120 g/L for women, 130 for men) would attribute the same risk for an individual with Hb of 119 g/L as for one with 53 g/L (the lowest in our material).

Third, to categorise the variable into categories that are clinically important, creating dummy variables. This approach could handle U-shaped risks. However, previously defined clinically important categories are needed and several degrees of freedom is spent in the analysis. As with a dichotomous transformation, all cases within a category are attributed the same risk.

d. To use a more complex fitting function, such as a restricted cubic spline [10, 11]. This approach uses so called knots, point estimates where the risk is determined. A cubic function is used to fit the function between knots. Near the ends the risk is modelled linear.

We prespecified Hemoglobin to be non-linear and tried the approaches above, see figure e4. We decided to use the 4-knot restricted cubic spline as both the best performance and was most appropriate from a clinical perspective. The knots were placed at the 5th, 35th, 65th and 95th percentiles where Hb was 92.25, 118, 130 and 148.15, respectively. The resulting function to fit Hb was:

```
2.2712251-0.017758194* hb-6.2295666e-06*pmax(hb-
92.25,0)^3+5.8240197e-05*pmax(hb-118,0)^3-7.7559735e-
05*pmax(hb-130,0)^3+2.5549104e-05*pmax(hb-148.15,0)^3
```

As opposed to the easily interpreted hazard ratio from the linear function, this is not easy to interpret without a graph, the graphic display of the four approaches is presented in figure e3.

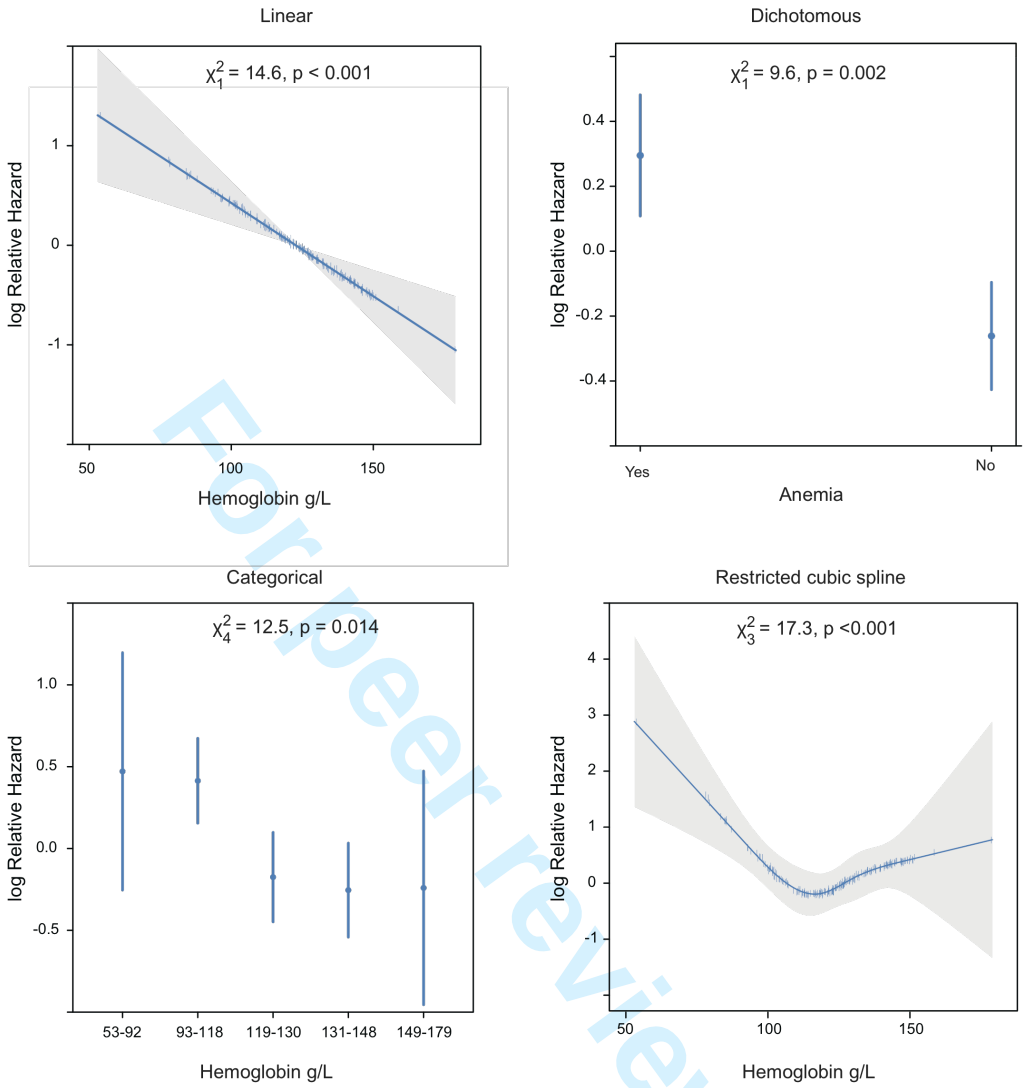

Figure e3. Different transformations of Hemoglobin. For dichotomous, the WHO definition of anemia is used. For categorical, the 5th, 35th, 65 and 95th percentiles were used, for easier comparison with the spline fit.

Apart from Hemoglogin, all other variables were bivariately tested for non-linearity by using 4-knot splines followed by ANOVA tests to determine if there was a significant non-linear component. GBS-ADL showed significant non-linearity and different codings were tested. We tested dichotomizing at the median and categorizing at the quartiles. A polynomial showed good fit but was not clinically plausible, with decreasing risks at the higher end of ADL impairment. The restricted cubic spline resembled a log fit and indeed the log fit was chosen, with fewer degrees of freedom spent, see figure e4. No other variables showed significant non-linear

effects.

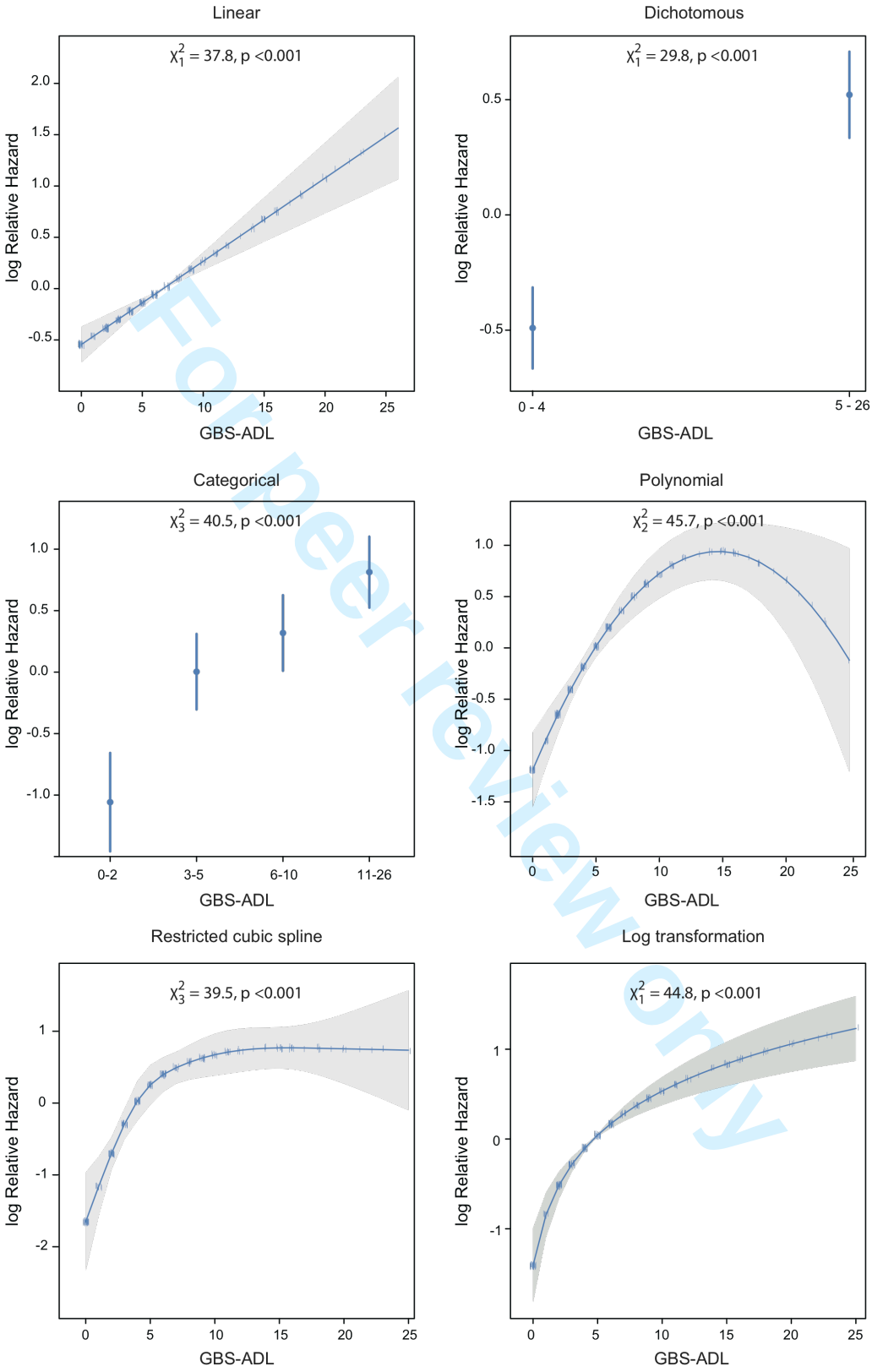

Figure e4. Different transformations tested for GBS-ADL.

5. Fitting the multivariate models

The two models were fitted, using the imputations and transformations above. The "model without ADL" used the covariates age, sex, charlson comorbidity index, albumin, BMI, eGFR, control/intervention status, and hemoglobin fitted as a restricted cubic spline The "full model" also included log(GBS-ADL).

6. Multivariate Diagnostics - Multicolinearity

Predictors with strong intercorrelations could cause interpreting problems, this is tested using the variance inflation factor (VIF). The interpretation of VIF has been disputed, a rule of thumb saying that VIF > 4 or > 10 signals a problematic multicolinearity problem have been suggested. However, these cut-offs may be too low, as a VIF over 10 could be acceptable [12]. To address multicolinearity, clustering of variables or data reduction could be applied.

In our models, all variables were simultaneously tested for colinearity. VIF Values were ranging between 1.02 and 1.47 in the "model without ADL" and between 1.10 and 1.52 in the "full model". The strongest bivariate correlation was between age and eGFR (r = -0.49). Thus, no apparent multicolinearity was present and no further action was taken.

7. Interactions – additivity assumption

A two-way interaction occurs when the effect of one predictor is dependending on the value of one other predictor. There are several recommendations regarding the number of interactions to test for. Only clinically plausible interactions could be tested, however, this requires prior knowledge. Another strategy is to test for all possible interactions, this requires a very large sample, to avoid overfitting. A compromise is to do a pooled interaction test for each variable and if the test is significant, the specific interactions are pursued [11].

In our study we did not have prespecified interactions for ADL and the sample size did not permit testing for all possible interactions. Therefore we opted for a global test approach. As we did not want to give ADL any advantages compared to the other variables, we also performed global tests for the other variables, one at a time. In the "model without ADL", the global test was significant for sex and BMI and an interaction term of sex * BMI was found (low BMI was a risk factor in men, not in women). This interaction was included in the model. In the "full model" another interaction, GBS-ADL*eGFR, was also found (the effect of impaired GBS-ADL was higher when eGFR was less imparied and vice versa). One interpretation of this interaction could be that impaired GBS-ADL is associated with weight loss and thus lower eGFR. To test properly for this we would need to apply three-way interactions (such as GBS-ADL*BMI*eGFR), which was beyond the scope of this paper.

8. Assumption of proportional hazards

The assumption of proportional hazards is the assumption that hazards from predictors do not vary over time. Proportional hazards can be tested in several different ways. Graphically, schoenfeld residuals are often plottet against time, then a straight line at zero is ideal. There are also different approaches to compensate for non-proportional hazards, the most common being adding an interaction term with time.

In our study, the PH assumption was first tested using a global test (`cox.zph` in R) as well as specific tests for all variables. In the "model without ADL", the global test gave a p value

of 0.72 and in the "full model" a p value of 0.70, signalling no violations of the PH assumption. The variable closest was eGFR, with a p value of 0.14. For eGFR, a schoenfeld residual plot is shown in figure e5. No further action was taken.

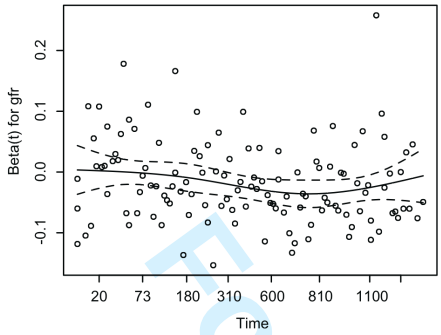

figure e5. Schoenfeld residual plot for eGFR.

9. Influential observations

With small sample size, a few influential observations could affect a model significantly. We screened for influential observations using dfBeta, that shows to what extent the regression coefficient would change, if that case should be removed. Every case is designated a dfBeta value for each variable. We used standardised dfBetas, with a cutoff of 0.20 to signify an influential observation. Thus, if deleting one observation led to a change in a predictor's β coefficient of more than 0.2 standard error, that observation was noted. As GBS-ADL was of specific interest, a sensitivity analysis was performed without the observations with dfBeta > 0.2 for ADL to determine whether the effect of GBS-ADL was only due to a few highly influential observations.

In the "model without ADL", a total of 23 (12%) observations had any DfBeta > ±0.20. The lowest dfBeta was -0.39 and the highest 0.32. In the "full model", 21 observations were considered influential. DfBetas ranged from -0.46 to 0.48. Nine cases had a dfBeta > ± 0.20 for GBS-ADL and/or its interaction with eGFR. A sensitivity analysis was done, with these nine observations excluded. In that model the overall χ^2 increased from 123 to 124 and the GBS-ADL χ^2 from 32 to 37. Thus, the effects of GBS-ADL in the "full model" were not due to a few influential observations. In all further analysis the influential observations were kept in the model.

10. Relative contribution of GBS-ADL

Describing the main effects of predictors including non-linear terms and interaction terms is not intuitive, especially not if the model contains a continuous-by-continuous interactions (such as eGFR * GBS-ADL). To obtain an estimate of the relative importance of the different predictors, we used the anova approach, developed by Harrell (`anova.rms` in R)[11]. Simple anova plots were included in the article as figure 1. Plots of the variable effects are shown below in figure e6. In these plots, interaction terms have been incorporated into the variables' relative importance.

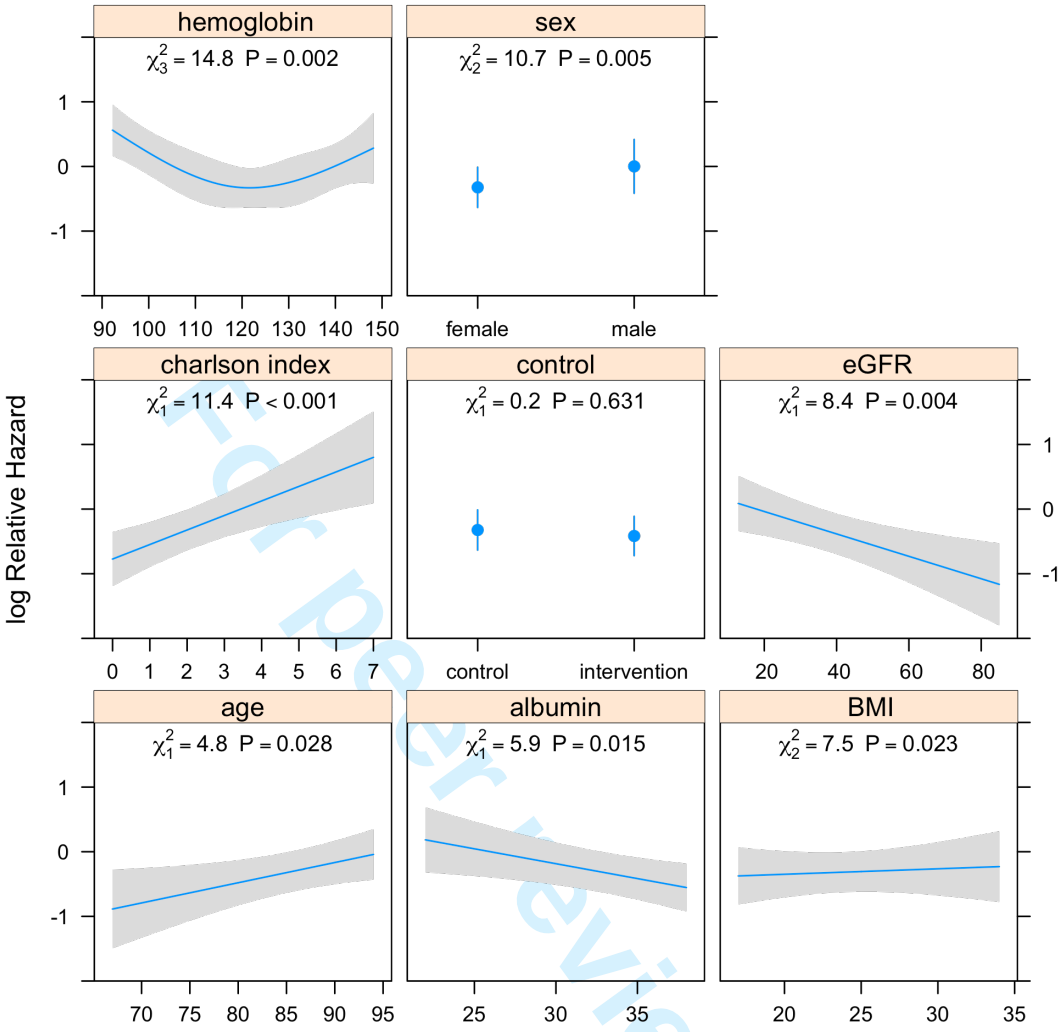

Figure e6a. Plot of variable effects in the “model without ADL”

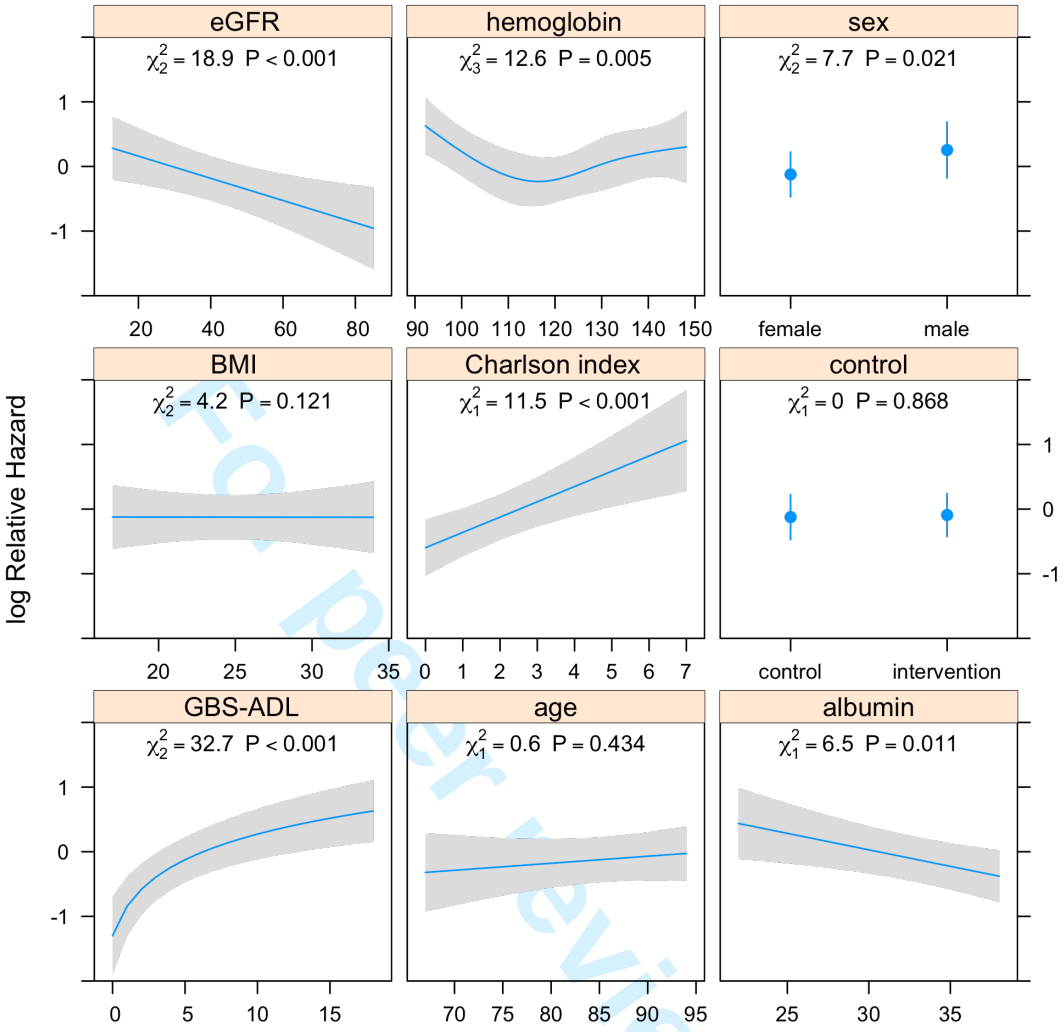

Figure e6b. Plots of the variable effects in the "full model".

11. Added value of GBS-ADL

There are several ways to determine the addded value of a variable in a regression model.

a. Likelihood ratio test. As we had two nested models (the "model without ADL" was also a part of the "full model") we performed a Likelihood ratio test as a χ^2 test over df = number of additional independent variables in the new model. The results are shown in table 3 in the article. For the "model without ADL", LR χ^2 was 78.4 and for the "full model" 121.0. The degrees of freedom were 11 and 13, respectively. Therefore the LR test resulted in a χ^2 (df = 2, N = 198) = 42.5, $p < 0.001$. Thus the "full model" had a significantly better fit.

b. Discrimination, measured with the C, or concordance, statistic. The C statistic is the probability that, in a case-control pair, the case (deceased) will be given a higher predicted risk from the model than the control (survivor). C statistics range from 0.5 (coin toss, useless) to 1.0 (perfect discrimination). In logistic regression (without time-to-event data), the c statistic is the same as ROC. For survival analysis, time is incorporated, so a case at time t is compared with a survivor at time t, albeit this survivor could be dead at time t+1 (the next

day). C statistics in survival analysis are often lower than ROC in logistic analysis. In addition, there are several different ways to calculate c statistic for time-to-event data.

We chose the method by Uno, to be able to compare between models. The "model without ADL" had a c statistic of 0.72 and the "full model" of 0.78. We set the follow-up time to 1428 days, as this was our median follow-up time of survivors. C statistics from the two models were compared using the method described by Uno et al. in the SurvC1 package [13]. Difference in c statistic between the model without ADL and the full model was 0.058 (95% CI = 0.022 - 0.094, p value 0.002).

c. NRI >0. Continuous net reclassification index[14, 15]. This index determines to what extent adding a new variable to a model leads to a change in the correct direction in predicted risk for each observation at time t (towards higher risk for deceased, towards lower for survivors). NRI>0 ranges from 0 (no increased value, useless) to 1(all observations reclassified in the right direction). NRI>0 has been shown to be more sensitive than change in C index, especially when the baseline model has a good performance. NRI>0 only describes the share of observations that have been reclassified, it does not quantify the amount of change in risk. Thus, it cannot distinguish between adding a variable that increases the predicted mortality risk for all cases with 1% or one that increases it with 50%.

For interpretation, the original NRI > 0 has been compared to the effect size of the added variable, where NRI>0 of 0.6 should be considered strong, 0.4 intermediate and below 0.2 weak [16]. However, after the initial development, Pencina et al. have suggested that ½ NRI>0 shoud be reported, as an average[15]. This is also what is given by the `IDI.INF` function in the `SurvIDINRI` package in R.

In our study ½ NRI>0 (95%CI) was 0.42 (0.22-0.58) with a p value <0.001. Again the follow-up time was set to 1428 days, to avoid extensive censoring. By doubling the point estimate of ½ NRI>0, the original NRI>0 would be 0.84, indicating a substantial effect size of adding ADL.

IDI. Integrated discrimination improvement. Originally developed by Pencina et al. for logistic models, IDI has been extended to time-to-event data [14, 17]. While NRI>0 displays the percentage of observtations being reclassified in the desirable direction, IDI is related to mean change in predicted probabilities within cases and controls. IDI is similar to testing the difference in R2, or discrimination slope, in logistic regrssion. IDI and NRI with confidence intervals were calculated (using the `IDI.INF` function) with the method by Uno et al. [18]. In our study IDI was 0.15 (95%CI 0.07-0.27, p < 0.001), indicating that the mean change in the correct direction was 15% (cases (deceased) were given 15% higher mortality risk by adding ADL while controls (survivors) were given 15% lower).

12. Overfitting - internal validation

Overfitting occurs when sample size is small. Then the fitted model becomes too optimistic and dependent on the present dataset. Thus, the findings will neither be reproducable nor valid in other populations. Ideally a model could be tested in another population at another location and setting, what is known as external validation. If that is not possible, there are several ways to accomplish internal validation. The recommended approach is via bootstrapping. In bootstrapping a new dataset, of the same size as the original, is constructed from the original dataset by resampling with replacement (an observation could be selected several times). This dataset is then used to develop the model, which is then tested on the full

original population. The difference in apparent performance and resampled performance is called optimism. The procedure is repeated 200-1000 times and the mean optimism is subtracted from the apparent performance estimates. This way an optimism-corrected estimate is obtained. In a future external population, this corrected estimate should be considered the best estimate possible[2, 8, 11, 19, 20].

Our aim was not to develop a valid prediction model, as this would have called for a larger sample size. Instead, we aimed to determine the relative imporance and added value of ADL when compared to the best possible traditional model. In the trade-off between overfitting and a well-performing traditional model, we empahsised the latter. A heuristic estimate of shrinkage would be χ^2 - d.f. / χ^2 = 123 - 14/123 = 0.89 for the "full model". However, this d.f. is falsely low as we tested many more interactions and transformations.

To better determine, the extent of overfitting, we carried out an internal validation with 1000 bootstraps. For the "model without ADL" and the "full model", the calibration slopes were 0.84 and 0.83, respectively, indicating a substantial amount of overfitting. The optimism-corrected R2 was 0.26 vs. 0.39. Optimism-corrected c statistics were 0.69 and 0.76.

13. Updating the model - final nomogram and Kaplan-Meier curves

Even if our aim was not to develop a valid prediction model, the amount of overfitting suggested that we should try to update the "full model". We chose to perform a LASSO (least absolute shrinkage and selection operator) model [21, 22]. In this model, the interaction terms and non-linear terms were combined into single terms. We used the `coxpath` function in R. Lasso could both shrink factors and eliminate variables. We considered the model with the lowest AIC (Akaike Information Criterion). In this model, the variable "control" was shrunk to zero, all other variables remained. The mean shrinkage was 0.84. The lasso path and AIC is shown in figure e7.

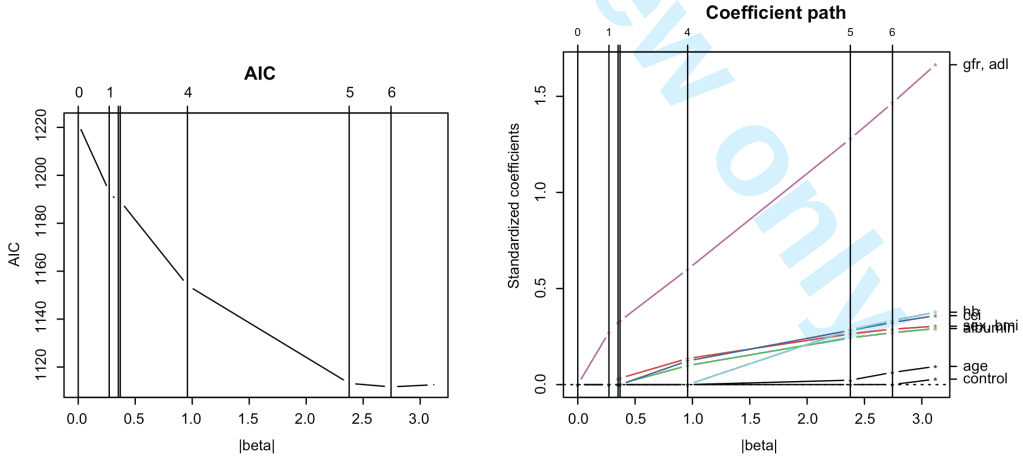

Figure e7. Lasso plots. AIC (Akaike information criterion) is lowest when control is set to zero.

A final nomogram, using the shrunk Lasso coefficients, was built, see figure e8. The cases were divided into four equally sized risk groups, by the quartiles of the linear predictor. To display the discrimination of the model, a kaplan-Meier curve was built, included in the article as figure 2.

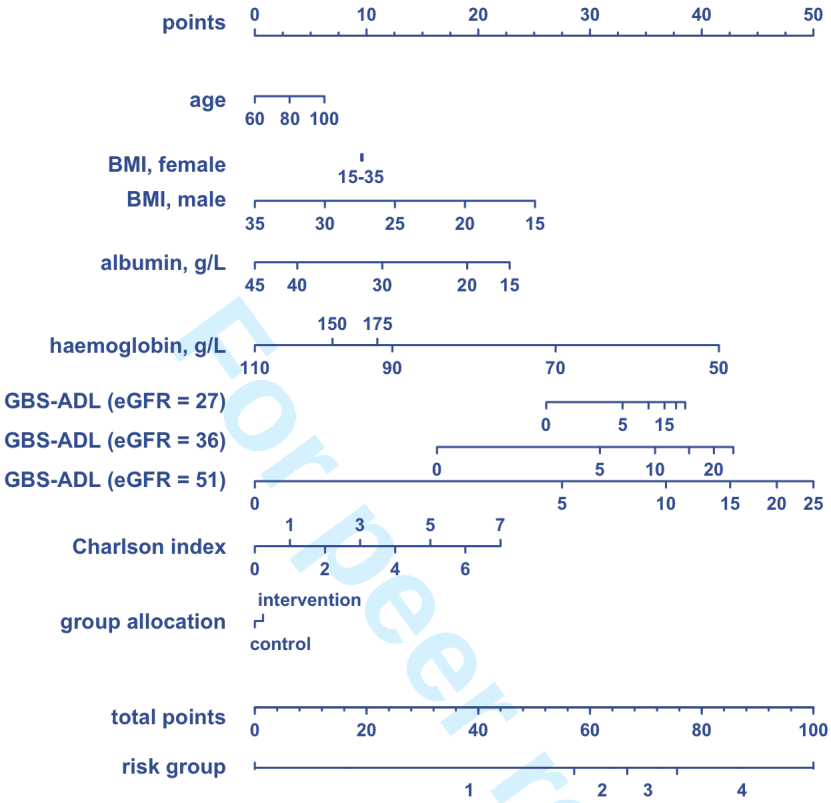

Figure e8a. Nomogram. Interpretation: For an individual, the variables are compared with the upper ”points” line, one at a time. These scores are then added for a total score that is plotted at the ”total points” line at the bottom. This could then be used to designate the person to a ”risk group” Notice the effect of interactions, low BMI is only a risk factor in men and the risk of GBS-ADL is moderated by eGFR, which is presented by median and quartiles. The cutoffs in the nomogram for the risk groups are completely arbitrary here, created to obtain 4 equally sized groups. In another scenario, cutoffs could be established to obtain for example a group with 90% chance of 3-year survival.

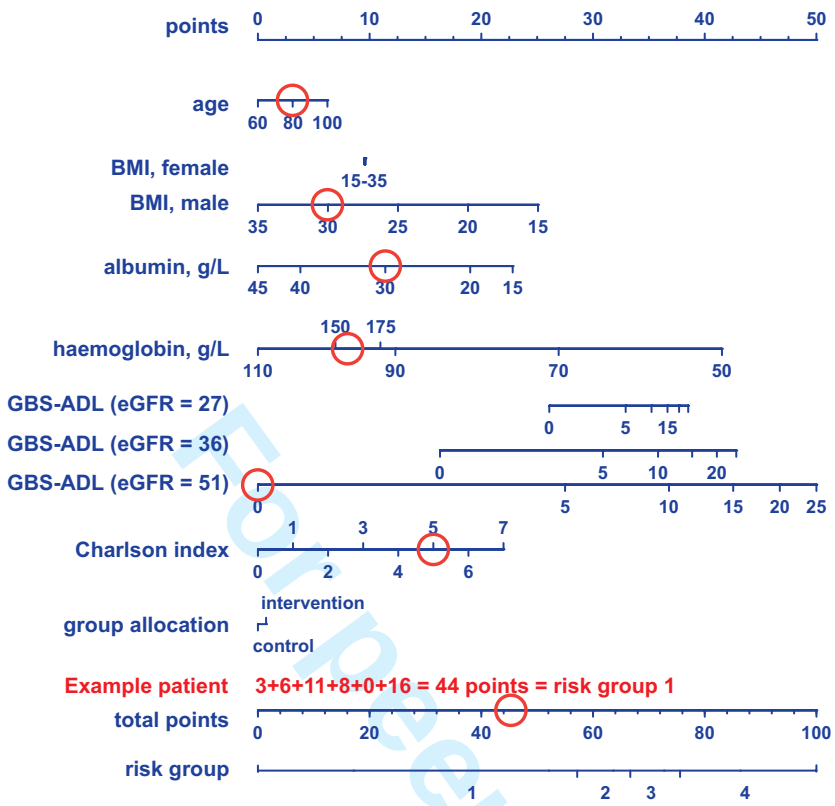

Figure e8b. Example of a scoring: This patient is 80 years old (3 points), male with BMI 30 (6 points), has an albumin of 30 (11 points) and a hemoglobin of 98 (8 points), normal kidney function and GBS-ADL (0 points) and a Charlson index score of 5 (16 points). The total score would be 3+6+11+8+0+16 = 44 points, placing this patient well within risk group 1. If this patient had all other variables constant but a functional decline, with a GBS-ADL score of 7, this would result in a total score of 44+30 = 74, placing the patient in risk group 3. The risk attributed to the functional decline would be equivalent to a hemoglobin drop from 98 to 55 g/L. Would it infer the same sense of urgency to the clinician?

References

1. *R: A language and environment for statistical computing.* [computer program]. Version. VIenna; 2014.
2. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* Feb 28 1996;15(4):361-387.
3. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol.* Oct 2006;59(10):1087-1091.
4. Horton NJ, Kleinman KP. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *Am Stat.* Feb 2007;61(1):79-90.
5. Janssen KJ, Donders AR, Harrell FE, Jr., et al. Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol.* Jul 2010;63(7):721-727.
6. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Med Res Methodol.* 2010;10:7.
7. Moons KG, Donders RA, Stijnen T, Harrell FE, Jr. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol.* Oct 2006;59(10):1092-1101.
8. Steyerberg EW. *Clinical Prediction Models - A Practical Approach to Development, Validation, and Updating*. New York: Springer; 2009.
9. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med.* Jan 15 2006;25(1):127-141.
10. Harrell FE, Jr., Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst.* Oct 5 1988;80(15):1198-1202.
11. Harrell Jr FE. *Regression Modeling Strategies - With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer-Verlag; 2001.
12. O'Brien RM. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity* 2007-10-01 2007;41(5):673-690.
13. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med.* May 10 2011;30(10):1105-1117.
14. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med.* Jan 30 2008;27(2):157-172; discussion 207-112.
15. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med.* Jan 15 2011;30(1):11-21.
16. Pencina MJ, D'Agostino RB, Pencina KM, Janssens AC, Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol.* Sep 15 2012;176(6):473-481.

17. Chambless LE, Cummiskey CP, Cui G. Several methods to assess improvement in risk prediction models: extension to survival analysis. *Stat Med.* Jan 15 2011;30(1):22-38.

18. Uno H, Tian L, Cai T, Kohane IS, Wei LJ. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Stat Med.* Jun 30 2013;32(14):2430-2442.

19. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med.* Aug 30 2004;23(16):2567-2586.

20. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* Aug 2001;54(8):774-781.

21. Tibshirani R. Regression shrinkage and selection via the Lasso. *J. Royal. Statist. Soc B.* 1996;58(1):267-288.

22. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med.* Feb 28 1997;16(4):385-395.

```
##########################################################
##                                                      ##
##            Appendix 2 - R code                       ##
##                                                      ##
##########################################################

## Data import from SPSS   ##

> library(rms)
> library(Hmisc)
> data <- spss.get() #file path in brackets

## Basic properties, table 1 (article) ##

> names(data)

"Nr"          "age"         "sex"         "status"
"time"        "cci"         "dressing"    "eating"
"physical"    "spontaneous" "hygiene"     "toilet"
"adl"         "hb"          "gfr"         "albumin"
"bnp"         "bmi"         "control"

> describe(data)

##########################################################
## 1. outcome                                           ##
##                                                      ##
##########################################################

## Determining follow-up and censoring ##

> library (survival)
> S <- Surv(time, status)
> cens.time <- ifelse(status == ”alive”, time, NA)
summary(cens.time)

## Baseline survival plot, figure 1 (appendix 1) ##

> S.years <- Surv(time/365.25, status)

> survplot(npsurv(S.years~1), xlab = "years", xlim =
c(0,4), time.inc = 1, lwd = 1.5, n.risk = T, y.n.risk =
0.05, cex.n.risk = 1, adj.n.risk = 0.5)

## discarding cases with missing outcome ##

> missing.outcome <- is.na(data$status)
> data <- data[missing.outcome == FALSE,]
> attach(data)

##########################################################
## 2. Crude analysis                                    ##
##                                                      ##
##########################################################
```

```
## Crude analysis for all predictors, table 2 (article)##

> summary(coxph(S~age))
> summary(coxph(S~sex))
> summary(coxph(S~cci))
> summary(coxph(S~dressing))
> summary(coxph(S~eating))
> summary(coxph(S~physical))
> summary(coxph(S~spontaneous))
> summary(coxph(S~hygiene))
> summary(coxph(S~toilet))
> summary(coxph(S~adl))
> summary(coxph(S~hb))
> summary(coxph(S~gfr))
> summary(coxph(S~albumin))
> summary(coxph(S~bnp))
> summary(coxph(S~bmi))
> summary(coxph(S~control))

## Discarding the separate ADL items ##

> data <- data[,-c(7:12)]

############################################################
## 3. Missing data                                        ##
##                                                        ##
############################################################

## Defining covariates ##

> covs <- data[, c("age", "sex", "cci", "adl", "hb",
"clearance", "albumin", "bnp", "bmi", ”control”)]

## Plotting missing, figure 2 (appendix 1) ##

> missing <- naclus(covs)
> naplot(missing, which = ”na per var”)

## Discarding the BNP variable ##

> data <- data[,-8]

## Determining associations with missing albumin ##

> missing.albumin <- ifelse(is.na(albumin), 1, 0)
> lrm(missing.albumin~age+sex+cci+clearance+bmi+adl+hb)
> chisq.test(missing.albumin, sex)
> oneway.test(cci~missing.albumin)
> tapply (cci, missing.albumin, mean)

## Creating a transcan object for imputation ##

> trans <-transcan(~age + sex + cci + adl + hb +
clearance + albumin + bmi + control, imputed = T, data =
data)
```

```
## Imputing albumin, gfr and bmi ##

> albumin.imputed <- as.integer(impute(trans, albumin,
data = data))
> gfr.imputed <- as.integer(impute(trans, gfr, data =
data))
> bmi.imputed <- as.integer(impute(trans, bmi, data =
data))

## Testing imputed variables, table 1 (appendix 1) ##

> summary(coxph(S~albumin.imputed)
> summary(coxph(S~gfr.imputed)
> summary(coxph(S~bmi.imputed)

## Imputed variables put into original dataset ##

> data$gfr <- gfr.imputed
> data$albumin <- albumin.imputed
> data$bmi <- bmi.1
> attach(data)

############################################################
## 4. Variable considerations                             ##
##                                                        ##
############################################################

## Screening for outliers, figure 3 (appendix 1) ##

> par(mfrow = c(3,3))
> boxplot(age)
> boxplot(cci)
> boxplot(hb)
> boxplot(gfr)
> boxplot(albumin)
> boxplot(bmi)
> boxplot(adl)

## Winsorising at 99th percentile, table 2(appendix 1) ##

> describe(gfr)
> gfr.winsorised <- ifelse(gfr > 118, 118, gfr)
> summary(coxph(S~gfr.winsorised))
> data$gfr <- gfr.winsorised
> attach(data)

############################################################
## Transformations for haemoglobin figure 4, appendix 1##
############################################################

## Range is obtained ##

> describe(hb)

## Linear model fitted and plotted ##
```

```
> hbfit.linear <- cph(S~hb, data = data)
> plot(Predict(hbfit.linear, hb = seq(53, 179, by = 1)),
xlab = "Haemoglobin g/L" , anova = anova(hbfit.linear),
pval = T, data = llist(hb))

## Dichotomous model##

> data$anemia <- ifelse(sex == ”male”, ifelse(hb < 130,
”yes”, ”no”), ifelse (hb < 120, ”yes”, ”no”))
> dd <- datadist (data)
> options(datadist = ”dd”)
> hbfit.dichotomous <- cph(S~anemia, data = data)
> plot(Predict(hbfit.dichotomous), xlab = "Anemia", anova
= anova(hbfit.dichotomous), pval = T)

## Categorical model ##

> data$hbcat <- ifelse(hb < 92.25, 1, ifelse(hb < 118, 2,
ifelse(hb < 130, 3, ifelse (hb < 148, 4, 5))))
> dd <- datadist (data)
> options(datadist = ”dd”)
> hbfit.categorical <- cph(S~as.factor(hbcat), data =
data)
> plot(Predict(hbfit.categorical), anova =
anova(hbfit.categorical), xlab = ”Haemoglobin g/L”, pval
= T)

## Restricted cubic spline ##

> hbfit.spline <- cph(S~rcs(hb,4), data = data)
> plot(Predict(hbfit.spline, hb = seq(53,179, by = 1)),
anova = anova(hbfit.spline), pval = T, xlab =
"Haemoglobin g/L", data = llist(hb))

## Testing other continuous variables ##

> agefit <- cph(S~rcs(age,4), data = data)
> anova(agefit)

> ccifit <- cph(S~rcs(cci,4), data = data)
> anova(ccifit)

> albuminfit <- cph(S~rcs(albumin,4), data = data)
> anova(albuminfit)

> bmifit <- cph(S~rcs(bmi,4), data = data)
> anova(bmifit)

> gfrfit <- cph(S~rcs(gfr,4), data = data)
> anova(gfrfit)

> adlfit <- cph(S~rcs(adl,4), data = data)
> anova(adlfit)
```

```
############################################################
## Transformations for ADL figure 5, appendix 1          ##
############################################################

> describe(adl)

## Linear fit ##

> adlfit.linear <- cph(S~adl, data = data)
> plot(Predict(adlfit.linear, adl = seq(0,25, by = 1)),
anova = anova(adlfit.linear), pval = T, data =
llist(adl), xlab = "GBS-ADL")

## Dichotomised at median ##

> data$adl.dichotomised <- ifelse(adl<5,0,1)
> dd <- datadist(data)
> options(datadist = "dd")
> adlfit.dichotomous <- cph(S~adl.dichotomised, data =
data)
> plot(Predict(adlfit.dichotomous), anova =
anova(adlfit.dichotomous), pval = T)

## Categorised at quartiles ##

> data$adl.quartiles <- ifelse(adl<3, 1, ifelse(adl<6,2,
ifelse(adl<10, 3,4)))
> dd <- datadist(data)
> options(datadist = "dd")
> adlfit.quartiles <- cph(S~as.factor(adl.quartiles),
data = data)
> plot(Predict(adlfit.quartiles), anova =
anova(adlfit.quartiles), pval = T)

## Two-degree polynomial ##

> adlfit.poly <- cph(S~pol(adl,2), data = data)
> plot(Predict(adlfit.poly, adl = seq(0,25,by=1)), anova
= anova(adlfit.poly), pval = T, xlab = "GBS-ADL", data =
llist(adl))

## four-knot spline ##

> adlfit.spline <- cph(S~rcs(adl,4), data = data)
> plot(Predict(adlfit.spline, adl = seq(0,25,by=1)),
anova = anova(adlfit.spline), pval = T, xlab = "GBS-ADL",
data = llist(adl))

## Log fit ##

> adlfit.log <- cph(S~log(adl+1), data = data)
> plot(Predict(adlfit.log, adl = seq(0,25,by=1)), anova =
anova(adlfit.log), pval = T, xlab = "GBS-ADL", data =
llist(adl))
```

```
#############################################################
## 5. Fitting the multivariate models                      ##
##                                                         ##
#############################################################

## The model without ADL ##

> model1 <- cph(S~age + sex + cci + rcs (hb, 4) + albumin
+ bmi + control + gfr, x = T, y = T, surv = T, data =
data)

## The full model ##

> model2 <- cph(S~age + sex + cci + rcs(hb, 4) + albumin
+ bmi + control + gfr + log (adl + 1), x = T, y = T, surv
= T, data = data)

#############################################################
## 6. Multicolinearity                                     ##
##                                                         ##
#############################################################

> cor(data)
> vif(model1)
> vif(model2)

#############################################################
## 7. Interactions                                         ##
##                                                         ##
#############################################################

## Global tests for the model without ADL ##

> z1 <- predict(model1, type = ”terms”)

> age.ia <- z1[,”age”]
> all.others <- z1[,-1]
> anova(cph(S~age.ia*all.others))

> sex.ia <- z1[,”sex”]
> all.others <- z1[,-2]
> anova(cph(S~sex.ia*all.others))
> cph(S~sex.ia*all.others)

> cci.ia <- z1[,"cci"]
> all.others <- z1[,-3]
> anova(cph(S~cci.ia*all.others))

> hb.ia <- z1[,"hb"]
> all.others <- z1[,-4]
> anova(cph(S~hb.ia*all.others))

> albumin.ia <- z1[,"albumin"]
> all.others <- z1[,-5]
> anova(cph(S~albumin.ia*all.others))
```

```
> bmi.ia <- z1[,"bmi"]
> all.others <- z1[,-6]
> anova(cph(S~bmi.ia*all.others))
> cph(S~bmi.ia*all.others)

> control.ia <- z1[,"control"]
> all.others <- z1[,-7]
> anova(cph(S~control.ia*all.others))

> gfr.ia <- z1[,"gfr"]
> all.others <- z1[,-8]
> anova(cph(S~gfr.ia*all.others))

## The full model ##

> z2 <- predict(model2, type = ”terms”)

> age.ia2 <- z2[,"age"]
> all.others.2 <- z2[,-1]
> anova(cph(S~age.ia2*all.others.2))

> sex.ia2 <- z2[,"sex"]
> all.others.2 <- z2[,-2]
> anova(cph(S~sex.ia2*all.others.2))

> cci.ia2 <- z2[,"cci"]
> all.others.2 <- z2[,-3]
> anova(cph(S~cci.ia2*all.others.2))

> hb.ia2 <- z2[,"hb"]
> all.others.2 <- z2[,-4]
> anova(cph(S~hb.ia2*all.others.2))

> albumin.ia2 <- z2[,"albumin"]
> all.others.2 <- z2[,-5]
> anova(cph(S~albumin.ia2*all.others.2))

> bmi.ia2 <- z2[,"bmi"]
> all.others.2 <- z2[,-6]
> anova(cph(S~bmi.ia2*all.others.2))

> control.ia2 <- z2[,"control"]
> all.others.2 <- z2[,-7]
> anova(cph(S~control.ia2*all.others.2))

> gfr.ia2 <- z2[,"gfr"]
> all.others.2 <- z2[,-8]
> anova(cph(S~gfr.ia2*all.others.2))
> cph(S~gfr.ia2*all.others.2)

> adl.ia <- z2[,"adl"]
> all.others <- z2[,-9]
> anova(cph(S~adl.ia*all.others))
> cph(S~adl.ia*all.others)
```

```
## Updating the models with the interactions ##

> model1 <- cph(S~age + sex * bmi + cci + rcs (hb, 4) +
albumin + control + gfr, data = data, x = T, y = T, surv
= T)

> model2 <- cph(S~age + sex * bmi + cci + rcs (hb, 4) +
albumin + control + gfr * log (adl + 1), data = data, x =
T, y = T, surv = T)

#########################################################
## 8. Proportional hazards assumption                   ##
##                                                      ##
#########################################################

> z3 <- predict(model1, type = "terms")
> model1.short <- cph(S~z3, x = T, y = T)
> ph1 <- cox.zph(model1.short, transform = "identity")
> ph1

> z4 <- predict(model2, type = "terms")
> model2.short <- cph(S~z4, x = T, y = T)
> ph2 <- cox.zph(model2.short, transform = "identity")
> ph2
> plot(ph2, var = "gfr") ##figure 6 (appendix 1) ##

#########################################################
## 9. Influential observations                          ##
##                                                      ##
#########################################################

> inf1 <- which.influence(model1)
> show.influence(inf1, dframe = data)

> inf2 <- which.influence(model2)
> show.influence(inf2, dframe = data)
> inf2

## Sensitivity analysis without influential for ADL ##

> subset <- data[-c(3,25,38,56,67,69,95,108,161),]
> attach(subset)
> S.sens <- Surv (time, status)
> sensitivity.model <- cph(S.sens~age + sex * bmi + cci+
rcs (hb, 4) + albumin + control + gfr * log(adl + 1), x=
T, y = T, surv = T, data = subset)
> sensitivity.model
> anova(sensitivity.model)
> detach(subset)
```

```
#############################################################
## 10. Relative contribution of ADL, figure 1(article) ##
##     figure 7(appendix)                                  ##
##                                                         ##
#############################################################

> plot(anova(model1), margin = "P", rm.ia = TRUE)
> plot(Predict(model1), anova = anova(model1), pval = T)

> plot(anova(model2), margin = "P", rm.ia = TRUE)
> plot(Predict(model2), anova = anova(model1), pval = T)

#############################################################
## 11. Added value of ADL, table 3 (article)             ##
##                                                         ##
#############################################################

## Likelihood ratio χ² test ##

> lrtest(model1, model2)

## Discrimination ##

> library (survC1)
> mydata <- as.matrix(data[,c("time", "status")])
> Inf.Cval.Delta(mydata, model1$x, model2$x, tau = 1428)

## NRI>0 and IDI ##

> library(survIDINRI)
> i <- IDI.INF(mydata, model1$x, model2$x, t0 = 1428)
> IDI.INF.OUT(i)

#############################################################
## 12. Internal validation                                ##
##                                                         ##
#############################################################

> validate(model1, B = 1000)
> validate(model2, B = 1000)

#############################################################
##                                                         ##
## 13. Updating the model                                 ##
##                                                         ##
#############################################################

>library(glmpath)
> mydata <- list(x = predict(model2, type = "ccterms"),
time = data$time, status = data$status)
> path <- coxpath(data = mydata)

##creating figure 8 appendix ##
```

```
> plot(path)
> plot(path, type = "aic")

## Determining the shrinkage factors ##

> lasso.factors <- path$b.predictor[path$aic ==
min(path$aic),]

## Shrinking the lasso.coefs ##

> lasso.coefs <- model2$coef

> lasso.coefs["age"] <- lasso.coefs["age"] *
lasso.factors[1]
> lasso.coefs["sex"] <- lasso.coefs["sex"] *
lasso.factors[2]
> lasso.coefs["bmi"] <- lasso.coefs["bmi"] *
lasso.factors[2]
> lasso.coefs["cci"] <- lasso.coefs["cci"] *
lasso.factors[3]
> lasso.coefs["hb"] <- lasso.coefs["hb"] *
lasso.factors[4]
> lasso.coefs["hb'"] <- lasso.coefs["hb'"] *
lasso.factors[4]
> lasso.coefs["hb''"] <- lasso.coefs["hb''"] *
lasso.factors[4]
> lasso.coefs["albumin"] <- lasso.coefs["albumin"] *
lasso.factors[5]
> lasso.coefs["control"] <- lasso.coefs["control"] * 0
> lasso.coefs["gfr"] <- lasso.coefs["gfr"] *
lasso.factors[7]
> lasso.coefs["adl"] <- lasso.coefs["adl"] *
lasso.factors[7]
> lasso.coefs["sex * bmi"] <- lasso.coefs["sex * bmi"] *
lasso.factors[2] * lasso.factors[2]
> lasso.coefs["gfr * adl"] <- lasso.coefs["gfr * adl"] *
lasso.factors[7] * lasso.factors[7]

## Updating the model ""

> lassomodel <- model2
> lassomodel$coefficients <- lasso.coefs

## Plotting nomogram, figure 2 (article) ##

> plot(nomogram(lassomodel, age = c(60,80,100), albumin =
c(15,20,30,40,45), bmi = c(15,20,25,30,35), hb =
c(50,70,90,110,150,175), interact = list(gfr =
c(27,36,51), adl, bmi, sex), lp = T, lp.at = c(-4,-
2,0,2), nint = 5, maxscale = 50))

## creating four risk groups ##
```

```
> risk.group <- cut2(as.numeric
(lassomodel$linear.predictor), g = 4)
> levels(risk.group) <- as.character(1:4)

## Kaplan-Meier plot, figure 3 (article)

> survplot(npsurv(S~risk.group, data = data), xlim =
c(0,1318), label.curves = FALSE, conf = "none", n.risk =
T, xlab = "follow-up (days)", cex.nrisk = 0.8, ylab =
"Fraction survivors", time.inc = 364, sep.n.risk = 0.03,
y.n.risk = 0, col = c(1,2,3,4), lty = 1)


### END ###
```

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract *"Prospective observation study" in title and abstract, page 1 and page 2*
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found *results section in abstract includes the most important findings, related to the objectives, page 2.*
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported *Rationale of improved mortality prediction [introduction, 1st paragraph] The need for improvement in studies regarding ADL and mortality [Introduction, 2nd paragraph*], page 4
Objectives	3	State specific objectives, including any prespecified hypotheses *"we aim to determine the relative importance and added value of this ADL measurement compared to clinical data, with regard to mortality prediction" [introduction 3rd paragraph], page 4*
Methods		
Study design	4	Present key elements of study design early in the paper *The methods section starts with a general description [Methods 1st paragraph], page 4. The1st paragraph of the statistical method section also describes the study design, page 6*
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection *setting, location, dates, follow-up and data collection are described on pages 4 to 6. Exposure was not dichotomous but continuous [Methods]*
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *eligibility, and methods for selection of participants [Methods, page 5] Follow-up [statistical method, page 7]*
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Not matched, Not applicable*
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable *all variables are presented in methods section [Methods, page 5 and 6] Variable transformations are presented in statistical methods section, page 7*
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group *sources of data and details of assessment are described. [Methods, page 5 and 6]*
Bias	9	Describe any efforts to address potential sources of bias *selection bias was addressed by including the group allocation variable in all analyses, as mentioned in discussion, page 13.*
Study size	10	Explain how the study size was arrived at

		"As a secondary analysis, no specific power calculation was done " [statistical methods, page 6]
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why *described in general in paper [statistical methods, page 6 to 9] and in detail in appendix 1*
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding *described in general in paper [statistical methods, page 6 to 9] and in detail in appendix 1*
		(*b*) Describe any methods used to examine subgroups and interactions *described in general in paper [statistical methods, page 7] and in detail in appendix 1*
		(*c*) Explain how missing data were addressed *described in general in paper [statistical methods, page 7] and in detail in appendix 1*
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *described in general in paper [results, page 9] and in detail in appendix 1*
		(*e*) Describe any sensitivity analyses *described in general in paper [statistical methods, page 7] and in detail in appendix 1*

Continued on next page

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed *eligibility is described in methods section and further details are found in the referenced original paper [methods page 5, reference 26]*
		(b) Give reasons for non-participation at each stage *This is described in the methods section and, in greater detail, in the referenced paper [methods page 5, reference 26]*
		(c) Consider use of a flow diagram *a reference to the flowchart in reference 26 is found in the methods section [methods page 5, reference 26]*
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders *descriptive data is found in table 1, page 9*
		(b) Indicate number of participants with missing data for each variable of interest *This is described in table 1, page 9 in the text of the results section, page 10 and in appendix 1.*
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount) *Follow-up time is summarised in results section, page 9 and appendix 1*
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time *number of events is reported in results section, page 9.*
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included *unadjusted estimates are presented in table 2 page 10. Multivariate estimates in figure 1, figure 2 and appendix 1.*
		(*b*) Report category boundaries when continuous variables were categorized *no categorisation or dichotomisation was done.*
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period *risks were displayed using a nomogram, figure 2*
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses *interactions are reported in statistical methods page 7, results page 11, appendix 1. One sensitivity analysis was done, with a summary in results page 11, larger detail in appendix 1.*
Discussion		
Key results	18	Summarise key results with reference to study objectives *[discussion 1st paragraph, page 12]*
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias *selection bias and Overfitting adressed in discussion page 12 and 13 and appendix 1 [methods].*
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence

		objectives, limitations, number of analysis and similar studies are discussed in discussion section, page 13.
Generalisability	21	Discuss the generalisability (external validity) of the study results *Validity (external and internal) is discussed in discussion section, page 13 and appendix 1.*
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based *Sources of funding, and roles of these, are presented in "funders", page 15.*

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Importance and added value of functional impairment in order to predict mortality: a cohort study in medical inpatients

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-014464.R1
Article Type:	Research
Date Submitted by the Author:	20-Jan-2017
Complete List of Authors:	Torisson, Gustav; Lunds Universitet, Clinical Memory Research Unit Dept of Clinical Sciences Stavenow, Lars; Skanes universitetssjukhus Malmo, Department of Internal Medicine Minthon, Lennart; Clinical Memory Research Unit, Department of Clinical Sciences Londos, Elisabet; Clinical Memory Research Unit, Department of Clinical Sciences, Lund University
Primary Subject Heading:	Geriatric medicine
Secondary Subject Heading:	Research methods
Keywords:	GERIATRIC MEDICINE, STATISTICS & RESEARCH METHODS, Delirium & cognitive disorders < PSYCHIATRY

Importance and added value of functional impairment in order to predict mortality: a cohort study in medical inpatients

Authors

Gustav Torisson[1], PhD, Gustav.Torisson@med.lu.se

Lars Stavenow[2], associate professor, Lars.E.Stavenow@skane.se

Lennart Minthon[1], professor, Lennart.Minthon@skane.se

Elisabet Londos[1], associate professor, Elisabet.Londos@skane.se

Affiliations

[1] Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden.

[2] Department of Internal Medicine, Skåne University Hospital, Malmö Sweden.

Corresponding author:

Gustav Torisson, Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden.

Adress: Simrisbanvägen 14, 20502 Malmö.

E-mail: gustav.torisson@med.lu.se.

Telephone: +46 40 33 77 85

Word count: 2857

Abstract

Background: Accurate estimation of prognosis in multimorbid hospital patients could improve quality of care. This study aims to determine the relative importance and added value of a performance-based ADL (activities of daily living) measure with regard to mortality prediction.

Methods: Two hundred inpatients, aged over 60 years, were recruited at the Department of General Internal Medicine at a tertiary university hospital. Two nested survival models were built, one with established risk factors (age, sex, Charlson comorbidity index, haemoglobin, albumin, body mass index, and glomerular filtration rate)**,** and one using the same covariates with the GBS (Gottfries-Bråne-Steen) ADL measure added. The relative importance of GBS-ADL was evaluated in the full model. The added value of GBS-ADL was determined by comparing the nested models using four approaches: difference in overall χ^2, discrimination, continuous net reclassification index (NRI > 0) and integrated discrimination improvement (IDI).

Results: In the full model, GBS-ADL was the single most important predictor of mortality (χ^2 - df = 30, p <0.001). The likelihood ratio χ^2 test showed significant added value of ADL (p<0.001). The c statistic was 0.78 with ADL and 0.72 without, (difference 0.058, 95% CI = 0.022 to 0.094). The NRI > 0 was 0.42 (95% CI 0.20 to 0.58) and IDI 0.15 (95% CI 0.07 to 0.22).

Conclusions: Compared to a set of available clinical risk factors, impairment in ADL was a stronger predictor of all-cause mortality, showing substantial added value. Implementing quantitative ADL measurements could enable more appropriate and individual care for the elderly.

Keywords: aging, comorbidity, mortality, functional status, statistical modeling

Strengths and limitations of this study

- A rigorous survival analysis was used to determine the relative importance of impaired ADL, compared to readily available clinical information.
- Four different methods were used to determine the added prognostic value of impaired ADL.
- However, the study was a secondary analysis, using data from an intervention study, and a larger study is needed.
- Only one ADL measurement was used, the results need to be confirmed for other ADL scales to be considered generalisable.

INTRODUCTION

Improving the accuracy of prognostic estimates could have several benefits for medical inpatients. Such benefits include reduced overtreatment, such as polypharmacy or the use of life-sustaining measures inconsistently with patients' preferences[1-4]. Other elderly patients are withheld treatment due to an incorrectly supposed poor prognosis, this could possibly be another important aspect[5-7]. Furthermore, patients with poor prognosis may prefer improved quality of life over extended survival. Therefore, accurate estimates could support doctors initiating a discussion regarding goals of care[8]. In addition, advance care planning could help patients and families to make necessary arrangements and increase quality of life[9-11].

Impairment in ADL (activities of daily living) is a well-known predictor of mortality and lower quality of life in hospitalized and community-dwelling elderly[12-20]. However, the majority of studies use interview-based scales[13, 15, 18], shown to differ significantly from performance-based ones[21, 22]. In addition, several studies use regression models without reporting overall performance[14, 15, 18, 23] and only few studies determine the added value of ADL[13, 15]. Recently, novel statistical methods have been introduced to establish the incremental value of prognostic markers[24].

In the present study, we aim to use these methods in order to determine the relative importance and added value of a performance-based ADL measure compared to clinical data, with regard to mortality prediction.

METHOD

This study constitutes a secondary analysis, all patients were concurrently taking part in a prospective trial, aiming to improve quality of care[25].

Setting

The study was carried out at the Department of General Internal Medicine at Skåne University Hospital in Malmö, Sweden. This teaching hospital provides care to the city's approximately 300,000 inhabitants. The department has four wards, with a total of 100 beds. Patients are admitted through the hospital's Emergency Department. Normally, the patients in the

department are elderly with multiple comorbidities. More specialized medical departments (cardiology, nephrology, endocrinology etc.) are separate and were not included in this study.

Patients

The recruitment of patients, that took place in 2009 and 2010, has been described in detail in a previous publication, including a flowchart[25]. In short, patients aged over 60 years, living in their own homes were eligible. Exclusion criteria comprised of terminal disease, language barrier, blindness/deafness/aphasia or other disease with inability to communicate, transfer to another department/ICU, early discharge and isolation due to communicable disease.

In total, 200 patients were included and underwent a baseline measurement. One half (101) of the patients constituted a control group while the other half (99) received a hospital-based, multidisciplinary intervention aiming to reduce rehospitalizations. The intervention consisted of a medication overview, improved discharge planning, telephone follow-up and improved liaison with GPs. Group allocation (intervention or control) used convenience sampling with geographic selection. At one-year follow-up, the intervention group had significantly fewer rehospitalizations than the control group[25].

ADL measurement

As part of the baseline measurement in the original trial, an ADL measurement was implemented by two experienced occupational therapists, who had received special training. The assessment was carried out when patients were stabilized, typically a few days into the admission.

The ADL subset of the GBS (Gottfries Bråne Steen) scale rates six items: dressing, food intake, physical activity, spontaneous activity, personal hygiene and toileting[26]. Items are scored on a performance-based 7-point scale ranging from 0 (best) to 6 (worst). For example, dressing is scored as follows:

0: Dresses and undresses without help

1:

2: Gets help with buttons, zips etc.

3:

4: Requires help from a caregiver to dress and undress but takes an active part

5:

6: Is completely dependent on a caregiver to be dressed and undressed

The points 1, 3 and 5 are not defined but are used by the observer to increase discrimination. Combining the six items gives a total ADL score of 0 (no impairment) to 36 (maximum impairment).

Other data from the original trial protocol

The Charlson comorbidity index was collected from the original protocol, to obtain a measure of combined comorbidity[27]. This index' performance concerning short-term and long-term mortality has recently been validated[28].

Data collection from medical records

Additional data was collected retrospectively regarding physiological and laboratory values. Since no blood samples were drawn in the original trial, only clinical data could be used. Candidate predictors were selected á priori on the basis of availability and previously established association with all-cause mortality. All data was obtained from the same hospital episode as ADL was measured. If unavailable during that hospitalization, the data point was labelled "missing". If several data points were found during the hospitalisation, the one closest to admission was used. The following variables, all independently related to all-cause mortality, were collected: Body mass index (BMI), kg/m^2, Hemoglobin (Hb), g/L, estimated Glomerular filtration rate (eGFR), ml/min, Albumin, g/L, Brain Natriuretic Peptide (BNP), ηg/L[29, 30, 31-33].

Statistical method

The present study was a secondary analysis, thus no specific power calculation was done beforehand, this had been done for the original intervention study, albeit with a different research question {Torisson, 2013 #215}. The goal of the present study was to compare the GBS-ADL measurement with the best set of available clinical risk factors using survival analysis. First, we built a multivariate Cox regression model, called "model without ADL", using the established risk factors as covariates. Then, this model was refitted, with ADL added, to obtain the "full model". To determine the added value of ADL, the performance of these two models were compared. In addition, the relative importance of ADL was examined in the "full model".

The modelling algorithm is based on previous recommendations, primarily by Harrell et al and Steyerberg et al [24, 34-37]. All steps are explained in larger detail in Supplementary File 1. All modelling was executed in R, the script is supplied in Supplementary File 2.

1. **Outcome**. The study endpoint was mortality status on Feb 6th 2014. Follow-up was defined as time from discharge of the original hospitalisation.

2. **Crude analysis**. Separate bivariate proportional hazards regressions were carried out for all variables on their original scaling. Crude analysis were accomplished for all separate ADL items but in further analysis only the total GBS-ADL score was used.

3. **Missing data.** Missing values in covariates were quantified and controlled for systematic patterns resulting in their missing status. Missing values were then imputed using an imputational regression model.

4. **Variable transformations**: Haemoglobin was pre-specified to have a non-linear association with mortality. All other continuous variables were tested for non-linearity and transformed accordingly. Outliers were controlled for data entry errors and considered for truncation.

5. **Fitting the two multivariate models.** The "model without ADL" was fitted first, using the transformations and imputations described above. Then, ADL was added and the model was refitted to obtain the "full model".

6. **Multicolinearity.** The models were tested using the VIF (variance inflation factor).

7. **Interactions**: Pooled two-way interaction tests were carried out for all variables, in both models, separately. If the pooled test was significant, specific interactions were pursued for that variable.

8. **Proportional hazards**. The proportional hazards assumption was tested with global tests and Schoenfeld residual plots for each variable.

9. **Influential observations.** Observations with a standardised DfBeta > 0.20 standard errors were noted for each variable. As ADL was of particular interest, a sensitivity analysis was performed without this variable's influential observations.

10. **Determining the relative importance of ADL.** As the models contained non-linear variables as well as interactions, simple measurements of main effects, such as hazard

ratios, could not be used. To obtain an estimate of the relative importance of the different predictors, an ANOVA test was used instead, where interaction terms and non-linear terms are incorporated into each variable.

11. **Determining added value of ADL**. To determine added value, the "model without ADL" and the "full model" were compared using:

    a. **Likelihood ratio test.** Performed as a χ^2 testing the difference in Likelihood ratio between the models' χ^2 over df = number of additional independent variables.

    b. **Discrimination**, measured with the C, or concordance, statistic. The C statistic is the probability that, in a case-control pair, the case will be given a higher predicted risk from the model than the control. C statistics ranges from 0.5 (coin toss, useless) to 1.0 (perfect discrimination). The difference in C statistic between models was tested using the method described by Uno et al.[38] .

    c. **NRI >0 (Continuous net reclassification index)**[39, 40]. This index determines to what extent adding a new variable leads to a change in the correct direction of predicted risk for each observation (towards higher risk for deceased, towards lower for survivors). NRI ranges from 0 (no increased value, useless) to 1(all cases reclassified in the right direction). NRI has been shown to be more sensitive than change in C index, especially when the baseline model has a good performance.

    d. **IDI**. Integrated discrimination improvement. Originally developed by Pencina et al. for logistic models, IDI has been extended to time-to-event data[39, 41]. While NRI>0 measures the percentage of observations that have been reclassified, it cannot distinguish between a small change in prediction and a large. IDI, however, measures the mean amount of such change. IDI and NRI with confidence intervals were calculated with the method by Uno et al.[42]

12. **Internal validation**. Both models were internally validated through 1000 bootstrap resamples to estimate the amount of overfitting and to obtain optimism-corrected performance estimates.

13. **Updating and presenting final model**. The "full model" was updated through the use of a LASSO (least absolute shrinkage and selection operator) procedure to reduce the effects of overfitting[43, 44]. The updated LASSO model was used to build a nomogram, with which patients were stratified into four equally sized risk groups, displayed in a Kaplan-Meier graph.

RESULTS

In two cases, mortality status could not be obtained; these were discarded from further analysis. Of the remaining 198 cases, 126 were deceased at follow-up. The median follow-up time for survivors was 1428 days (range 1312-1548). Baseline characteristics are displayed in table 1.

Table 1. Baseline characteristics

Continuous variables	**mean (SD)**	**median (IQR)**	**min-max**
Age	83.4 (8.1)	85 (78-89)	60-100
Charlson comorbidity index	2.3 (1.5)	2 (1-3)	0-7
GBS-ADL, total	6.8 (5.7)	5 (2-10)	0-25
GBS-ADL, dressing	1.3 (1.4)	1(0-2)	0-5
GBS-ADL, food intake	0.1 (0.4)	0(0-0)	0-2
GBS-ADL, physical activity	2.0 (1.1)	2(2-2)	0-5
GBS-ADL, spontaneous activity	1.0 (1.2)	1(0-2)	0-5
GBS-ADL, hygiene	1.4 (1.4)	2(0-2)	0-5
GBS.ADL, toilet	0.9 (1.4)	0(0-1)	0-6
Hemoglobin, g/L	123 (19)	124 (111-136)	53-179
eGFR, ml/min, n = 197	42.3 (25)	37(26-51)	6-198
BMI, kg/m2, n = 195	24.7 (5.1)	24 (21-27)	14-42
Albumin, g/L, n = 181	31.5 (4.9)	32 (29-35)	14-42
BNP, ηg/L, n = 85	261 (297)	147 (54-377)	3-1618
Categorical variables	**number**	**percentage**	
Male sex	70	35%	
In intervention group in original study	99	50%	

Table 1. Baseline characteristics for the entire sample. n = 200 unless otherwise stated. ADL = activities of daily living, eGFR =estimated glomerular filtration rate, BMI = Body mass index, BNP = Brain natriuretic peptide.

The results from the crude analysis are presented in table 2.

Table 2. Crude analysis

Predictor	β	S.E	Wald X^2	p value	HR (95% CI)
GBS-ADL-total, points	0.08	0.013	37.8	<0.001	1.08 (1.06 - 1.11)
GBS-ADL-hygiene, points	0.38	0.06	37.7	<0.001	1.46 (1.29 - 1.65)
GBS-ADL-physical, points	0.46	0.08	36.0	<0.001	1.59 (1.36 - 1.84)
GBS-ADL-dressing, points	0.31	0.06	30.0	<0.001	1.36 (1.22 - 1.52)
eGFR, ml/min, n = 197	-0.029	0.005	29.3	<0.001	0.97 (0.96 - 0.98)
GBS-ADL-spontaneous, points	0.33	0.06	27.0	<0.001	1.40 (1.23 - 1.58)
Charlson index, points	0.22	0.06	15.2	<0.001	1.25 (1.18 - 1.40)
Hemoglobin, g/L	-0.019	0.005	14.6	<0.001	0.98 (0.97 - 0.99)
Albumin, g/L, n = 181	-0.064	0.018	13.1	<0.001	0.94 (0.90 - 0.97)
GBS-ADL- toileting, points	0.19	0.05	11.6	<0.001	1.20 (1.08 - 1.34)
Age, years	0.036	0.011	10.1	0.001	1.04 (1.01 - 1.06)
BMI, kg/m^2 , n = 195	-0.053	0.020	7.4	0.007	0.95 (0.91 - 0.99)
BNP, ηg/L, n = 85	0.0009	0.0003	6.7	0.01	1.001 (1 - 1.002)
ADL - food intake, points	0.34	0.21	2.7	0.10	1.41 (0.93 - 2.12)
Sex (0 = Female, 1 = Male)	0.29	0.18	2.6	0.11	1.34 (0.94 - 1.92)
Group in original study (0=control, 1=intervention)	0.11	0.18	0.37	0.54	1.12 (0.78 - 1.59)

Table 2. Crude Cox regression for all predictors, sorted by decreasing strength of association. S.E = standard error, HR = Hazard ratio, ADL = Activities of daily living, eGFR = glomerular filtration rate, BMI = body mass index, BNP = brain natriuretic peptide.

BNP was missing in 115 cases (58%) and the variable was discarded from further analysis. eGFR and BMI were missing in 1 and 3 cases, respectively; these were considered to be missing completely at random. Albumin was missing in 17 cases, these were predominantly female (15/17) and had lower scores on Charlson comorbidity index. Missing values were imputed with a minimal change in variable properties, see Supplementary File 1.

Hemoglobin was fitted using a 4-knot restricted spline and GBS-ADL was transformed using the natural logarithm. No other predictors showed significant non-linear properties and they were kept in their original form. eGFR had one extreme outlier at 198ml/min that was winsorized at the 99th percentile (118 ml//min).

A significant sex * BMI interaction was found and included into the models (low BMI was a significant predictor in men but not in women). Another interaction, eGFR * ADL, was included as well (ADL was a stronger predictor when eGFR was unimpaired and vice versa). No other significant interactions were found. No significant multicolinearities were found. The proportional hazards assumption was not violated. In the full model, 21 observations were influential, of which 9 for ADL and/or its interaction with eGFR. A sensitivity analysis with these cases removed showed a slight improvement in model fit and is presented in Supplementary File 1. However, all observations were kept in the models.

In the "full model", ADL was by far the most significant predictor. The relative importance of the predictor variables are shown in figure 1. All four measurements showed added value for model with ADL, see table 3.

Table 3. Added value of ADL

Model comparison	model without ADL	model with ADL	p value
Nagelkerke R^2	0.33	0.46	
Likelihood ratio χ^2	78.4 (11df)	121.0(13df)	<0.001
c statistic (95% CI)	0.72(0.67-0.76)	0.78(0.73-0.82)	0.001
½ NRI > 0 (95% CI)		0.42(0.20-0.58)	<0.001
IDI (95% CI)		0.15(0.07-0.22)	<0.001

Table 3. Comparison of the two nested survival models. NRI > 0 = continuous Net Reclassification Index, IDI = Integrated Discriminatory Improvement.

When bootstrapped 1000 times, the calibration slope of the "model without ADL" was 0.84 and of the "full model" 0.83. Optimism-corrected R^2 was 0.27 vs. 0.40, respectively. Optimism-corrected c statistics were 0.70 and 0.76. When the LASSO was employed to shrink coefficients and update the model, the mean shrinkage was 0.84. The nomogram using the updated model coefficients is shown in Supplementary File 1 and the subsequent Kaplan-Meier graph for the four risk groups are presented in figure 2.

DISCUSSION

In this study, we confirm that impaired ADL is an important predictor of mortality in elderly medical inpatients. The relative contribution of ADL was larger than of the available predictors in a real-life setting, including a comorbidity index, available physiological parameters and laboratory values. In addition, ADL showed a substantial added value when compared to a model combining all of these traditional predictors.

In the crude analysis, four of the GBS-ADL items were stronger predictors than the Charlson comorbidity index. Thus, a simple rating of dressing ability had better predictive value than a combined comorbidity measure, designed to predict mortality. This indicates that performance-based ADL measures are truly important mortality predictors in multimorbid patients. In multivariate analysis, impairment in ADL was by far the most important predictor and all four measures signaled added value when GBS-ADL was added to the traditional predictors.

The mechanism underlying the association between ADL and mortality is probably multifactorial. Impairment in ADL could contribute directly to mortality in some aspects. Obvious complications to functional decline include pressure sores, atrophy, falls, thrombosis etc. However, less intuitive factors could also apply, such as attaining multi-resistant bacteria or Clostridium Difficile[45 46]. Even more likely, ADL acts a proxy for a confounder not measured by the model. A possible such confounder is frailty, defined as an increased vulnerability, where small stressors lead to adverse outcomes, such as hospitalization or death[47]. The frailty phenotype includes unintentional weight loss, along with loss of strength, low physical activity, slow walking speed and exhaustion[48]. There is a considerable overlap between frailty, comorbidity and ADL impairment. Our study utilized specific measures for comorbidity and ADL impairment, but not for frailty. However, our model is most likely describing the effects of frailty as well.

Several methodological issues need to be addressed. First, the choice of ADL scale, where the GBS scale was chosen in order to facilitate implementation locally. There are large variations and lack of standardization regarding functional measures used in medical inpatients[49]. The GBS scale proved feasible and has been shown to have a good construct validity and inter-rater reliability[50]. In addition, the GBS-ADL has correlated strongly with other ADL measurements, for example Katz' index[51, 52]. Ideally, two different scales should have

been employed, to enable a comparison between scales and possibly improve generalisability. A potentially confounding issue was the concurrent non-randomized trial. However, the variable "control/intervention status" was included in all analyses, without any sign of bias. In addition, no power calculation was done, the sample size was small, and internal validation showed that our models were indeed overfitted, with a calibration slope of 0.83. This overfitting is probably not a result of having too many covariates but rather a result of the global interaction tests and tests of non-linearity. This multiple comparison situation has been called "testimation bias"[37]. The overall aim was not to develop the most comprehensive and parsimonious prediction model to use in future populations but to describe the importance and added value of ADL. Therefore, we prioritized not to miss clinically important interactions and/or transformations in the trade-off with overfitting. To compensate partly, we used a LASSO procedure to shrink estimates. The small sample size and the aim to compare ADL with the best possible model was also the reason underlying the imputation of missing values. In addition, the main diagnosis of the current hospitalisation was not included as a predictor in the analysis. The reason for this was the large heterogeneity of main diagnoses (with 97 different ICD codes in 200 patients), albeit this could possibly have been achieved with a larger sample size as well.

The primary strength of this study is the rigorous statistical approach. State-of-the-art methods were used in the model building to handle missing data, to address non-linearity, to screen for interactions, for model diagnostics and for internal validation. In addition, four different methods were applied to estimate added value. Previously, a study has showed increase in model χ^2 when adding a composite ADL measure, regarding 2-year mortality [15]. However, this study compared ADL only with comorbidity indices. With such a limited reference model it is likely that a new measure will add value but the final model could still perform poorly, which was reflected by low model χ^2 values and a final c statistic of 0.66. The use of comorbidity indices only as reference model is also far from the clinical reality. Another study shows increase in discrimination when adding an ADL measurement to a 1-year logistic regression mortality prediction model[13]. This study also starts with comorbidity indices alone and does not report any other measurement of overall performance (such as overall χ^2 or R^2). Our study compares ADL to a much more complex reference model and yet shows added value using both these previously applied measurements as well as several others.

Implications for further research include research regarding performance-based ADL scales, including the relation to specific frailty ratings. Larger studies could obtain head-to-head comparisons of ADL vs. disease-specific predictors, such as ejection fraction in heart failure.

Today, ADL is very often assessed in a variety of ways in medical inpatients, to assess the individuals' needs after discharge. Implementing a performance-based quantitative measurement could have many benefits, also apart from prognostic value, such as increased standardization and the possibility to follow a patient over time. As a final remark, mortality prediction is not all about avoiding overtreatment due to a poor prognosis. Our model identified 50 elderly multimorbid medical inpatients with a 90% chance of 3-year survival. This group should not be undertreated simply due to age discrimination.

In conclusion, an ADL measurement showed significant added value as a predictor of mortality in a multimorbid elderly hospital population. Implementation of standardized ADL measurements could lead to better prognostic estimates and in the end a more appropriate and individualised care for the elderly.

DECLARATIONS

List of abbreviations

ADL: activities of daily living

GBS: Gottfries-Bråne-Steen

BMI: body mass index

eGFR: estimated glomerular filtration rate

BNP: brain natriuretic peptide

NRI: net reclassification index

IDI: integrated discrimination improvement

Ethics approval

All patients enrolled in the original study gave written informed consent. Both the original trial and the secondary analysis have been approved by the regional ethics committee at Lund University.

Availability of data and materials

Since the participatns were not specifically asked for consent to share data, such sharing is not compatible with the current Swedish legislation. The data protection officer at Skåne University hospital, the data protection officer at Lund university as well as lawyers at the Swedish data protection authority have unanimously adviced us not to publish data, even if anonymized.

Conflicts of interests

The authors declare no financial relationships with any organizations that might have an interest in the submitted work and no other relationships or activities that could appear to have influenced the submitted work.

Funders

This study was financed by the Swedish Research Council (Vetenskapsrådet #523-2010-520), the Swedish Brain Power programme, the National Swedish Board of Health and Welfare and the Governmental Funding of Clinical Research within the National Health Services (ALF). All researchers acted independently to the funding bodies. The funding agencies had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Authors' contributions

The study was conceived by LS and LM. Data collection was done by GT, supervised by EL. GT performed all analysis and drafted the manuscript, which was critically revised by LS, LM and EL. All authors have approved the final version.

Acknowledgments

The authors would like to thand Sofia Raccuia, Jenny Cappelin and Anna Johansson for their assistance in the data acquisition.

REFERENCES

1. Frankl D, Oye RK, Bellamy PE. Attitudes of hospitalized patients toward life support: a survey of 200 medical inpatients. The American journal of medicine 1989;**86**(6):645-8
2. Heyland DK, Barwich D, Pichora D, et al. Failure to engage hospitalized elderly patients and their families in advance care planning. JAMA internal medicine 2013;**173**(9):778-87.
3. Somogyi-Zalud E, Zhong Z, Hamel MB, et al. The use of life-sustaining treatments in hospitalized persons aged 80 and older. Journal of the American Geriatrics Society 2002;**50**(5):930-4
4. Teno JM, Gozalo PL, Bynum JP, et al. Change in end-of-life care for Medicare beneficiaries: site of death, place of care, and health care transitions in 2000, 2005, and 2009. JAMA : the journal of the American Medical Association 2013;**309**(5):470-7.
5. Biganzoli L, Wildiers H, Oakman C, et al. Management of elderly patients with breast cancer: updated recommendations of the International Society of Geriatric Oncology (SIOG) and European Society of Breast Cancer Specialists (EUSOMA). The Lancet Oncology 2012;**13**(4):e148-60.
6. Tahir M, Robinson T, Stotter A. How not to neglect the care of elderly breast cancer patients? Breast 2011;**20**(4):293-6.
7. Wildiers H, Kunkler I, Biganzoli L, et al. Management of breast cancer in elderly individuals: recommendations of the International Society of Geriatric Oncology. The Lancet Oncology 2007;**8**(12):1101-15.
8. Dev S, Clare RM, Felker GM, et al. Link between decisions regarding resuscitation and preferences for quality over length of life with heart failure. European journal of heart failure 2012;**14**(1):45-53.
9. Steinhauser KE, Christakis NA, Clipp EC, et al. Factors considered important at the end of life by patients, family, physicians, and other care providers. JAMA : the journal of the American Medical Association 2000;**284**(19):2476-82
10. Detering KM, Hancock AD, Reade MC, et al. The impact of advance care planning on end of life care in elderly patients: randomised controlled trial. BMJ 2010;**340**:c1345.
11. Wright AA, Zhang B, Ray A, et al. Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. JAMA : the journal of the American Medical Association 2008;**300**(14):1665-73.
12. Han PK, Lee M, Reeve BB, et al. Development of a prognostic model for six-month mortality in older adults with declining health. Journal of pain and symptom management 2012;**43**(3):527-39.
13. Covinsky KE, Justice AC, Rosenthal GE, et al. Measuring prognosis and case mix in hospitalized elders. The importance of functional status. Journal of general internal medicine 1997;**12**(4):203-8
14. Fried LP, Kronmal RA, Newman AB, et al. Risk factors for 5-year mortality in older adults: the Cardiovascular Health Study. JAMA : the journal of the American Medical Association 1998;**279**(8):585-92
15. Inouye SK, Peduzzi PN, Robison JT, et al. Importance of functional measures in predicting mortality among older hospitalized patients. JAMA : the journal of the American Medical Association 1998;**279**(15):1187-93
16. Lee SJ, Lindquist K, Segal MR, et al. Development and validation of a prognostic index for 4-year mortality in older adults. JAMA : the journal of the American Medical Association 2006;**295**(7):801-8.

17. Teno JM, Harrell FE, Jr., Knaus W, et al. Prediction of survival for older hospitalized patients: the HELP survival model. Hospitalized Elderly Longitudinal Project. Journal of the American Geriatrics Society 2000;**48**(5 Suppl):S16-24
18. Walter LC, Brand RJ, Counsell SR, et al. Development and validation of a prognostic index for 1-year mortality in older adults after hospitalization. JAMA : the journal of the American Medical Association 2001;**285**(23):2987-94
19. Torisson G, Minthon L, Stavenow L, et al. Cognitive impairment is undetected in medical inpatients: a study of mortality and recognition amongst healthcare professionals. BMC geriatrics 2012;**12**:47.
20. Torisson G, Stavenow L, Minthon L, et al. Reliability, validity and clinical correlates of the Quality of Life in Alzheimer's disease (QoL-AD) scale in medical inpatients. Health and quality of life outcomes 2016;**14**:90.
21. Magaziner J, Zimmerman SI, Gruber-Baldini AL, et al. Proxy reporting in five areas of functional status. Comparison with self-reports and observations of performance. American journal of epidemiology 1997;**146**(5):418-28
22. Sherman SE, Reuben D. Measures of functional status in community-dwelling elders. Journal of general internal medicine 1998;**13**(12):817-23
23. Reuben DB, Rubenstein LV, Hirsch SH, et al. Value of functional status as a predictor of mortality: results of a prospective study. The American journal of medicine 1992;**93**(6):663-9
24. Steyerberg EW, Pencina MJ, Lingsma HF, et al. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. European journal of clinical investigation 2012;**42**(2):216-28.
25. Torisson G, Minthon L, Stavenow L, et al. Multidisciplinary intervention reducing readmissions in medical inpatients: a prospective, non-randomized study. Clinical interventions in aging 2013;**8**:1295-304.
26. Gottfries CG, Brane G, Gullberg B, et al. A new rating scale for dementia syndromes. Archives of gerontology and geriatrics 1982;**1**(4):311-30
27. Charlson M, Szatrowski TP, Peterson J, et al. Validation of a combined comorbidity index. Journal of clinical epidemiology 1994;**47**(11):1245-51
28. Frenkel WJ, Jongerius EJ, Mandjes-van Uitert MJ, et al. Validation of the Charlson Comorbidity Index in acutely hospitalized elderly adults: a prospective cohort study. Journal of the American Geriatrics Society 2014;**62**(2):342-6.
29. Janssen I, Katzmarzyk PT, Ross R. Body mass index is inversely related to mortality in older people after adjustment for waist circumference. Journal of the American Geriatrics Society 2005;**53**(12):2112-8.
30. Culleton BF, Manns BJ, Zhang J, et al. Impact of anemia on hospitalization and mortality in older adults. Blood 2006;**107**(10):3841-6.
31. Tonelli M, Wiebe N, Culleton B, et al. Chronic kidney disease and mortality risk: a systematic review. Journal of the American Society of Nephrology : JASN 2006;**17**(7):2034-47.
32. Goldwasser P, Feldman J. Association of serum albumin and mortality risk. Journal of clinical epidemiology 1997;**50**(6):693-703
33. Wallen T, Landahl S, Hedner T, et al. Brain natriuretic peptide predicts mortality in the elderly. Heart 1997;**77**(3):264-7
34. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine 1996;**15**(4):361-87.

35. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. European heart journal 2014;**35**(29):1925-31.
36. Harrell Jr FE. *Regression Modeling Strategies - With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer-Verlag, 2001.
37. Steyerberg EW. *Clinical Prediction Models - A Practical Approach to Development, Validation, and Updating*. New York: Springer, 2009.
38. Uno H, Cai T, Pencina MJ, et al. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Statistics in medicine 2011;**30**(10):1105-17.
39. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Statistics in medicine 2008;**27**(2):157-72; discussion 207-12.
40. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Statistics in medicine 2011;**30**(1):11-21.
41. Chambless LE, Cummiskey CP, Cui G. Several methods to assess improvement in risk prediction models: extension to survival analysis. Statistics in medicine 2011;**30**(1):22-38.
42. Uno H, Tian L, Cai T, et al. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. Statistics in medicine 2013;**32**(14):2430-42.
43. Tibshirani R. Regression shrinkage and selection via the Lasso. J Royal Statist Soc B 1996;**58**(1):267-88
44. Tibshirani R. The lasso method for variable selection in the Cox model. Statistics in medicine 1997;**16**(4):385-95
45. Chen TY, Anderson DJ, Chopra T, et al. Poor functional status is an independent predictor of surgical site infections due to methicillin-resistant Staphylococcus aureus in older adults. Journal of the American Geriatrics Society 2010;**58**(3):527-32.
46. Rao K, Micic D, Chenoweth E, et al. Poor functional status as a risk factor for severe Clostridium difficile infection in hospitalized older adults. Journal of the American Geriatrics Society 2013;**61**(10):1738-42.
47. Clegg A, Young J, Iliffe S, et al. Frailty in elderly people. Lancet 2013;**381**(9868):752-62.
48. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. The journals of gerontology Series A, Biological sciences and medical sciences 2001;**56**(3):M146-56
49. Buurman BM, van Munster BC, Korevaar JC, et al. Variability in measuring (instrumental) activities of daily living functioning and functional decline in hospitalized older medical patients: a systematic review. Journal of clinical epidemiology 2011;**64**(6):619-27.
50. Brane G, Gottfries CG, Winblad B. The Gottfries-Brane-Steen scale: validity, reliability and application in anti-dementia drug trials. Dementia and geriatric cognitive disorders 2001;**12**(1):1-14
51. Dencker K, Gottfries CG. Activities of daily living ratings of elderly people using Katz' ADL Index and the GBS-M scale. Scandinavian journal of caring sciences 1995;**9**(1):35-40
52. Katz S, Ford AB, Moskowitz RW, et al. Studies of Illness in the Aged. The Index of Adl: A Standardized Measure of Biological and Psychosocial Function. JAMA : the journal of the American Medical Association 1963;**185**:914-9

Figure Legends

Figure 1. Relative importance of predictors in the multivariate “model without ADL” and the “full model”. Interaction terms and non-linear effects have been incorporated in the variables. A higher χ^2-df value indicates a stronger association. Control = the grouping variable from the original study. BMI = body mass index, eGFR = estimated glomerular filtration rate.

Figure 2. Kaplan-Meier estimates from the updated full model including customary risk factors and ADL. The participants have been stratified into four equally sized groups by quartiles of risk.

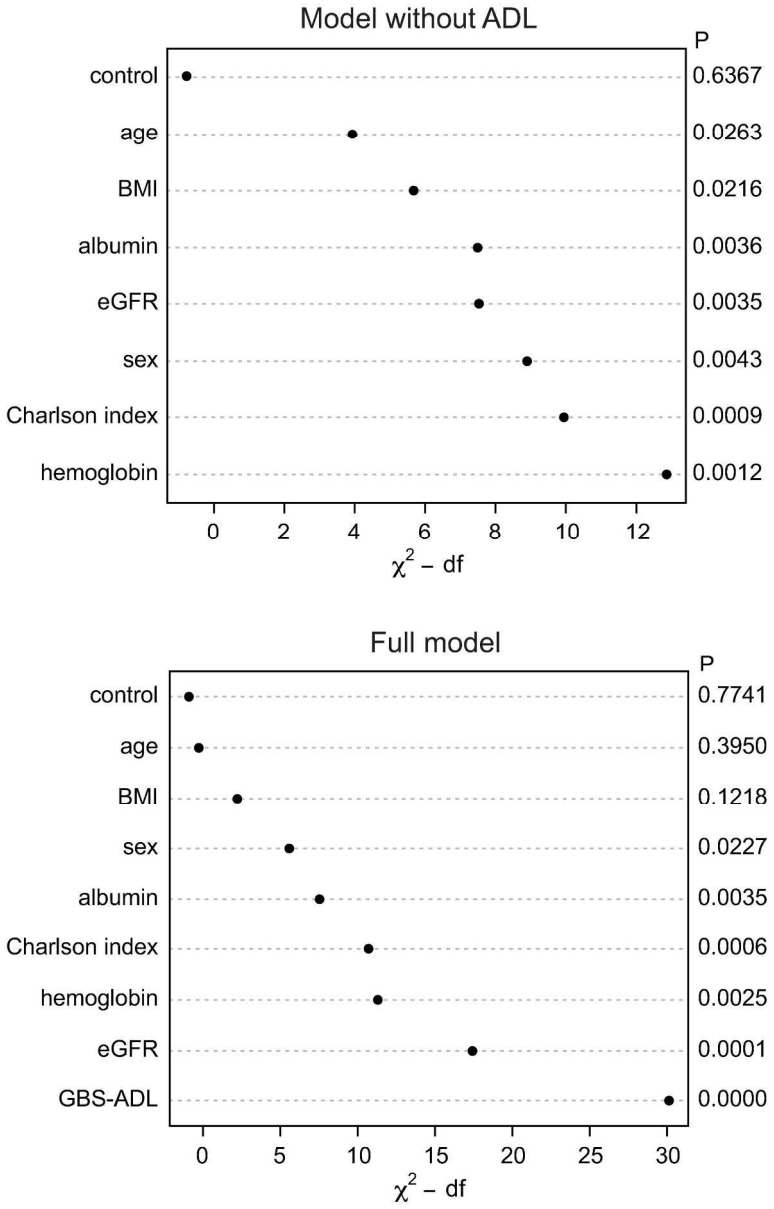

Figure 1. Relative importance of predictors in the multivariate "model without ADL" and the "full model". Interaction terms and non-linear effects have been incorporated in the variables. A higher χ2-df value indicates a stronger association. Control = the grouping variable from the original study. BMI = body mass index, eGFR = estimated glomerular filtration rate.

189x298mm (300 x 300 DPI)

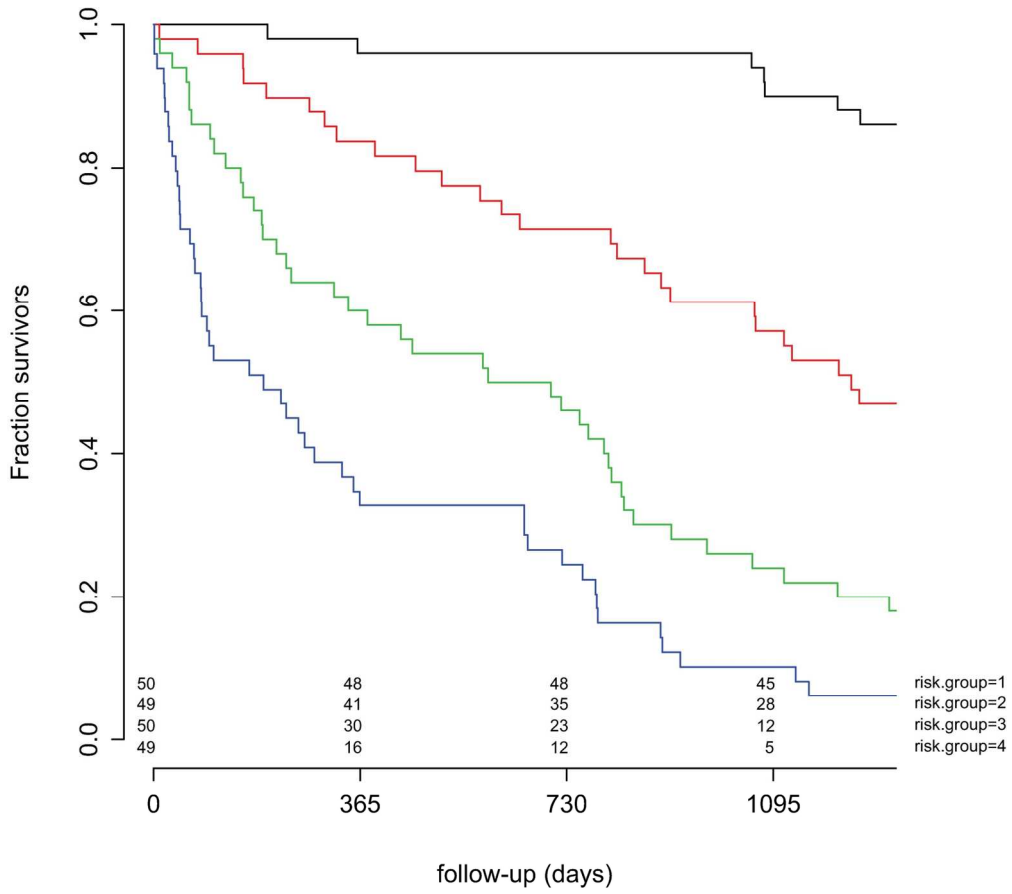

Figure 2. Kaplan-Meier estimates from the updated full model including customary risk factors and ADL. The participants have been stratified into four equally sized groups by quartiles of risk.

149x132mm (300 x 300 DPI)

Appendix 1 –statistics

General aspects

The overall aim was to compare ADL with the best possible model containing customary risk factors.
To achieve this, two models were developed and compared, the "model without ADL" and the "full model".
Data was originally stored in an SPSS file. All data analysis was performed in R [1]. Code is provided in appendix 2.

1. Outcome

Generally, it is important to describe the quantity of cases with missing outcome and to determine if there are any underlying patterns. Otherwise, simple exclusion may affect representativity [2].

In the study

Survival status was determined using the local region's electronic registry on the 6th February 2014. The time variable was defined as days from discharge to death or censoring at study endpoint, whichever came first. Those surviving at endpoint had been followed for a median of 1428 days (range 1312-1548). The baseline survival function is shown in figure e1.

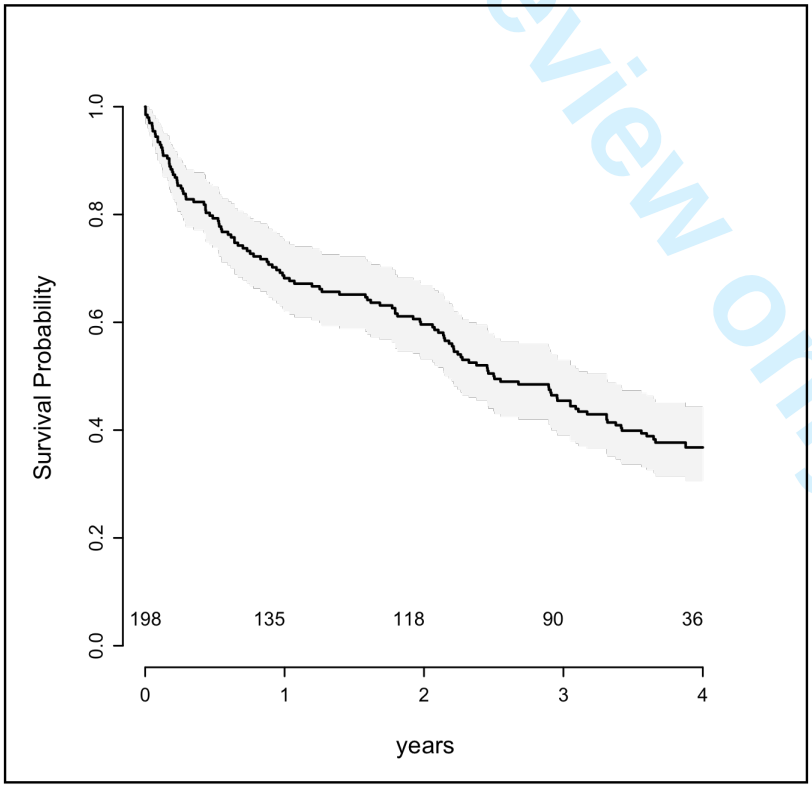

figure e1. Baseline survival function.

In our study, two cases were missing survival status due to having moved abroad (no longer in the region's registry) Hypothetically, these cases could be assumed to be in better health (severely diseased patients are unlikely to move abroad?). However, they were considered too few to affect representativity and were discarded from further analysis. Thus, the number of cases decreased from 200 to 198.

2. Crude analysis

Before any modifications are done to a variable, a crude analysis for the intented outcome could be of interest, to obtain an initial estimate of the effect of the predictor

In the study
Bivariate Cox proportional hazards regressions were carried out separately for all variables, including only outcome and the variable. All variables were treated in their original form, on their original scale. Observations with missing values were excluded from crude analysis. Data is presented with β coefficients, Standard errors, Wald χ^2, p value and hazard ratios in table 2 in the article.

In the crude analysis, all variables/potential predictors were statistically significant except sex, control/intervention status in the original study and the ADL item "food intake". Regarding the latter, the distribution was severely skewed, with only 18 cases (9%) having a non-zero value. To obtain a preliminary ranking of importance, the variables were sorted by decreasing Wald χ^2 in the table in the article.

In crude analysis, all separate GBS-ADL items were included but in further multivariate analysis, only the total GBS-ADL score was used, to avoid fitting too many variables and multicolinearity (the ADL items were intercorrelated at r = 0.8-0.9)

3. Missing data

In general, it is important to analyse missing data patterns in predictors. The first step is to determine the quantity of missing data. The second step is whether data is missing completely at random or if there are underlying patterns. When these prerequisites have been fulfilled, there are several approaches to missing data:

1. Listwise deletion, discarding all observations with any missing data points. The advantage of this approach is that no "manipulation" is done. Therefore, this method may seem intuitively most correct. The obvious disadvantage is that sample size could be substantially diminished. In addition, representativity could be affected, if missing a variable is systematically associated with other characteristics.

2. Using simple imputation. This technique substitutes missing values with the mean, mode or median value. This could be acceptable only if the variable is missing completely at random and the percentage of missing values small.

3. Using a more complex imputational technique. This approach uses customised regression models including all other covariates to obtain a stable prediction of the missing values. This method has been described and emphasized in several publications [3-7]. When using complex imputations, single or multiple imputations could be chosen. In the latter case, a separate

dataset is analysed for each imputational iteration, leading to a much larger complexity in the analysis.

In the study

When analysing the quantity of missing data, eGFR was missing in one case, BMI in three cases. Albumin was missing in 17 (9%) cases. BNP was missing in 113 (56%) observations.

The BNP variable was discarded from further analysis, as it had more than 50% missing. BMI and eGFR were considered missing completely at random. However, we found that cases with missing albumin were predominantly female (15 female vs. 2 male, $\chi^2 = 3.32$, $p = 0.056$) and had lower score on Charlson comorbidity index (1.47 vs. 2.33, $F = 11.3$, $p = 0.002$). Thus, excluding cases with missing albumin would affect representativity. Discarding the albumin variable would affect the overall aim, to compare ADL with the best possible traditional model. Therefore, the missing values in BMI, eGFR and Albumin were imputed using a single conditional imputation method (with the `transcan` function in R). In total, the effect of imputations was very small on the variable properties, as shown below.

Variable	β	S.E	Wald X2	p value	HR (95% CI)
Albumin, g/L, n = 181	-0.064	0.018	13.1	<0.001	0.94 (0.90 - 0.97)
- with imputation, transcan	-0.066	0.017	14.7	<0.001	0.94 (0.91 - 0.97)
eGFR, ml/min, n = 197	-0.029	0.005	29.3	<0.001	0.97 (0.96 - 0.98)
- with imputation, transcan	-0.029	0.005	29.2	<0.001	0.97 (0.96 - 0.98)
BMI, kg/m^2 , n = 195	-0.053	0.020	7.4	0.007	0.95 (0.91 - 0.99)
- with imputation, transcan	-0.053	0.020	7.4	0.006	0.95 (0.91 - 0.99)

Table e1. Effect of imputation on variable properties.

4. Variable considerations

Extreme outliers

In regression, outliers may be defined as observations with more than 3 interquartile ranges over the third quartile or below the 1st quartile. Such extreme values may affect a regression model significantly. First data entry errors should be considered and pursued. Then the biological plausibility should be considered. If plausible, we may consider a truncation at the 99th or 1st percentile [8].

In the study

In our study, data screening revealed, that for eGFR there was one extreme outlier with an estimated value of 198 ml /min (> 6 IQR over 3rd quartile), see boxplot.

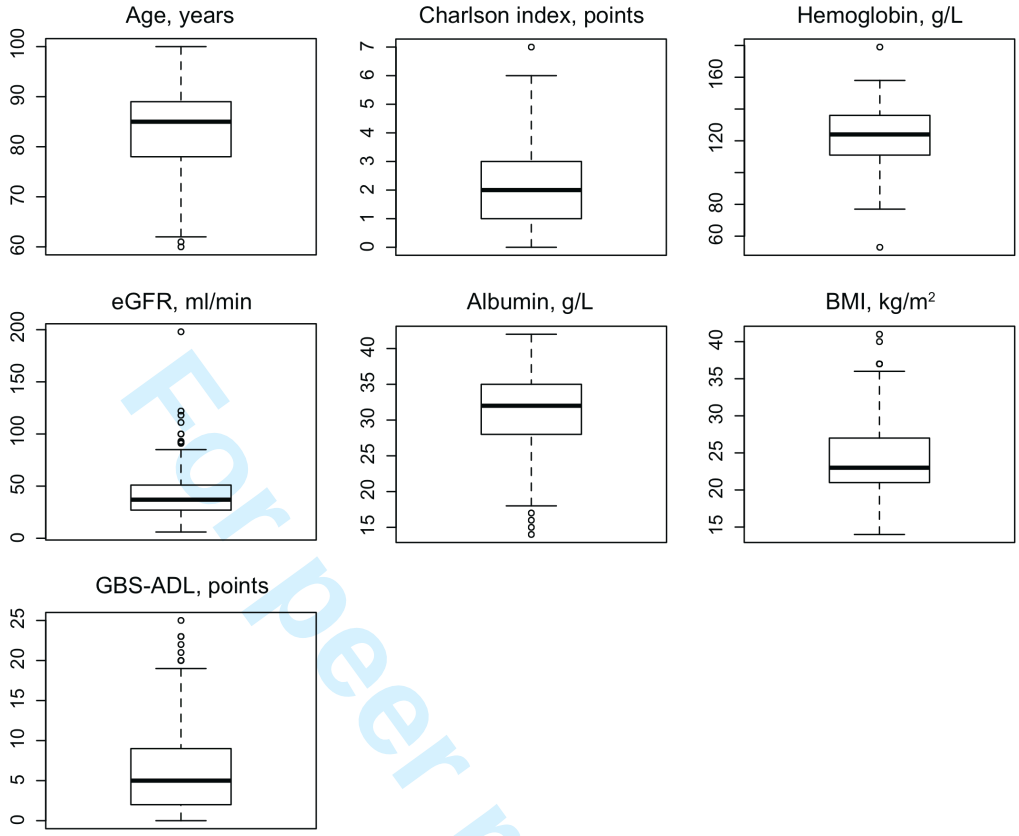

Figure e2. Boxplots of the continuous predictors. eGFR = Glomerular filtration rate, BMI = Body mass index, ADL = Activities of daily living.

This case was screened for data entry errors but none were found. Regarding biological plausibility, eGFR was measured with the Cockcroft-Gault formula ((140-age) * weight * constant)/Serum Creatinine in μmol/L, where the constant is 1.23 for men and 1.04 for women. Thus, GFR was not measured directly, but estimated and sensitive to extreme values in both serum creatinine, age and body weight. With this reservation, we considered the value to be biologically plausible. However, we did not consider it clinically important to compare one elderly patient with 198 ml/min in eGFR with another with 120 ml/min with regard to mortality. Therefore, eGFR was winsorized at the 99th percentile (118 ml/min). This led to a slightly improved fit in univariate performance.

Variable	β	S.E	Wald X2	p value	HR (95% CI)
eGFR, ml/min	-0.029	0.005	29.2	<0.001	0.97 (0.96 - 0.98)
- winsorized at 99th percentile	-0.029	0.005	29.8	<0.001	0.97 (0.96 - 0.98)

Table e2. effect of winsorization on variable properties.

Non-linearity

Most regression model assume that the predictors are linearly related to the outcome. However, non-linear relationships, such as U-shapes, for continuous variables are common.

There are several ways to address non-linearity:

First, assuming that the variable is linear. The advantage of this approach is that it results in an easily interpreted main effect, for example the Hazard Ratio in survival analysis. This is

the approach used in our crude comparisons. However, the approach is potentially problematic. For hemoglobin, this would mean that the risk difference between two individuals with 170 and 130 g/L would be the same as between two with 90 and 50, respectively. In addition, this approach cannot handle U-shaped risks, it is likely that someone with 200 g/L in Haemoglobin with dehydration or polycytemia does not have better survival than someone with 140 g/L

Second, to dichotomize the variable, using a previously established cut-off, is another frequently used approach. However it is not recommended as it ignores a lot of information [9]. In our example, applying the WHO cut-off for anemia (120 g/L for women, 130 for men) would attribute the same risk for an individual with Hb of 119 g/L as for one with 53 g/L (the lowest in our material).

Third, to categorise the variable into categories that are clinically important, creating dummy variables. This approach could handle U-shaped risks. However, previously defined clinically important categories are needed and several degrees of freedom is spent in the analysis. As with a dichotomous transformation, all cases within a category are attributed the same risk.

Fourth, to use a more complex fitting function, such as a restricted cubic spline [10, 11]. This approach uses so called knots, point estimates where the risk is determined. A cubic function is used to fit the function between knots. Near the ends the risk is modelled linear.

In the study

We prespecified Hemoglobin to be non-linear and tried the approaches above, see figure e4. We decided to use the 4-knot restricted cubic spline as both the best performance and was most appropriate from a clinical perspective. The knots were placed at the 5th, 35th, 65th and 95th percentiles where Hb was 92.25, 118, 130 and 148.15, respectively. The resulting function to fit Hb was:

```
2.2712251-0.017758194* hb-6.2295666e-06*pmax(hb-
92.25,0)^3+5.8240197e-05*pmax(hb-118,0)^3-7.7559735e-
05*pmax(hb-130,0)^3+2.5549104e-05*pmax(hb-148.15,0)^3
```

As opposed to the easily interpreted hazard ratio from the linear function, this is not easy to interpret without a graph, the graphic display of the four approaches is presented in figure e3.

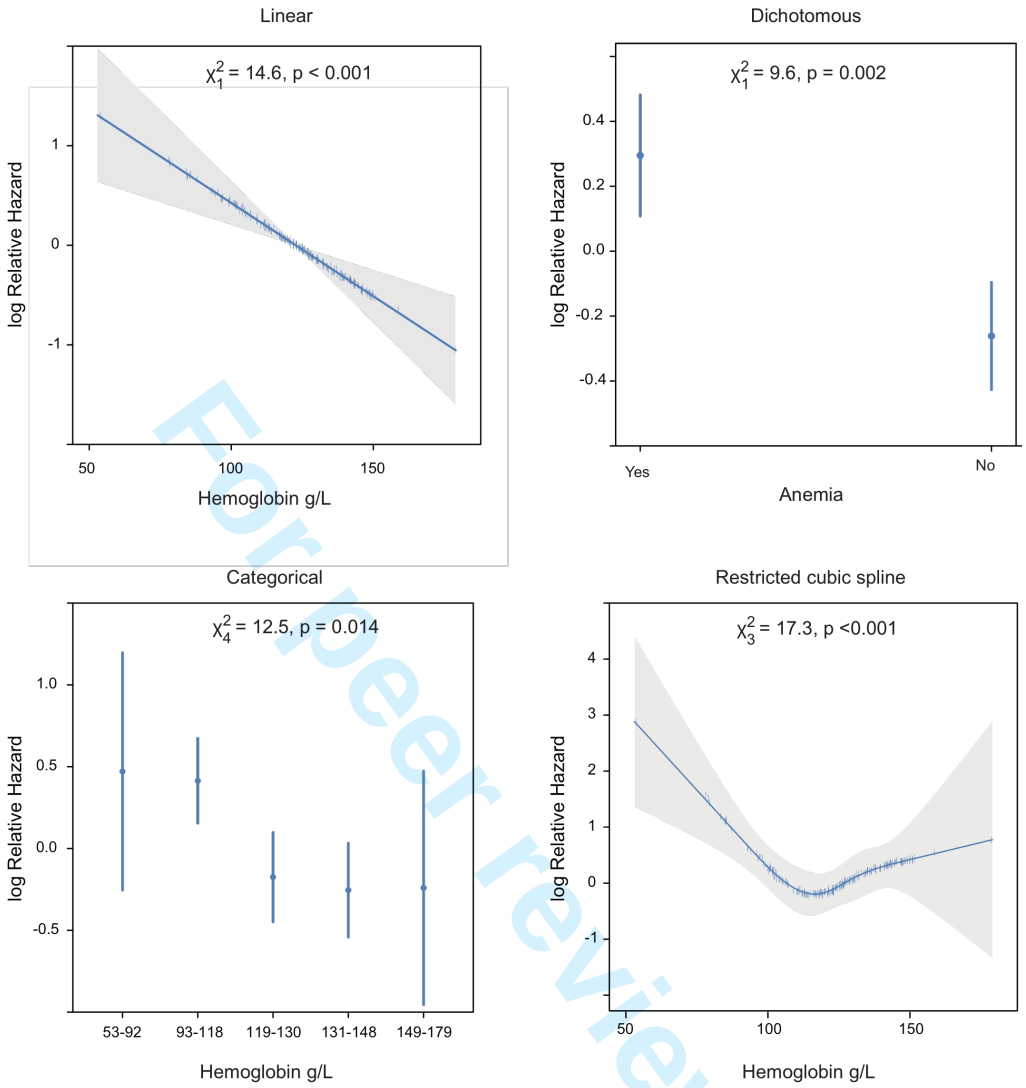

Figure e3. Different transformations of Hemoglobin. For dichotomous, the WHO definition of anemia is used. For categorical, the 5th, 35th, 65 and 95th percentiles were used, for easier comparison with the spline fit.

Apart from Hemoglogin, all other variables were bivariately tested for non-linearity by using 4-knot splines followed by ANOVA tests to determine if there was a significant non-linear component. GBS-ADL showed significant non-linearity and different codings were tested. We tested dichotomizing at the median and categorizing at the quartiles. A polynomial showed good fit but was not clinically plausible, with decreasing risks at the higher end of ADL impairment. The restricted cubic spline resembled a log fit and indeed the log fit was chosen, with fewer degrees of freedom spent, see figure e4. No other variables showed significant non-linear effects.

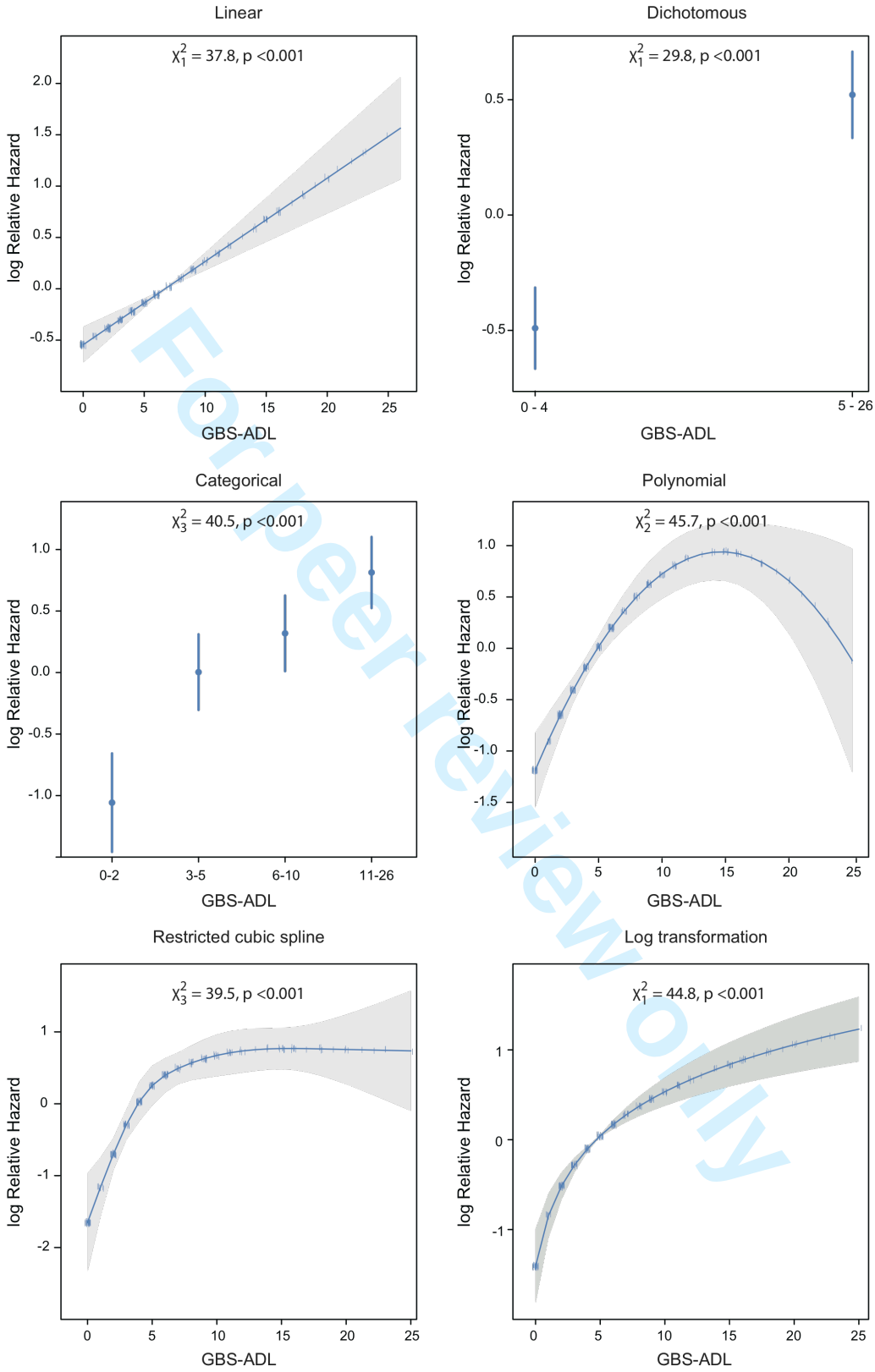

Figure e4. Different transformations tested for GBS-ADL.

5. Fitting the multivariate models

In the study
The two models were fitted, using the imputations and transformations above. The "model without ADL" used the covariates age, sex, charlson comorbidity index, albumin, BMI, eGFR, control/intervention status, and hemoglobin fitted as a restricted cubic spline The "full model" also included log(GBS-ADL).

6. Multivariate Diagnostics - Multicolinearity

Predictors with strong intercorrelations could cause interpreting problems, this is tested using the variance inflation factor (VIF). The interpretation of VIF has been disputed, a rule of thumb saying that VIF > 4 or > 10 signals a problematic multicolinearity problem have been suggested. However, these cut-offs may be too low, as a VIF over 10 could be acceptable [12]. To address multicolinearity, clustering of variables or data reduction could be applied.

In the study
In our models, all variables were simultaneously tested for colinearity. VIF Values were ranging between 1.02 and 1.47 in the "model without ADL" and between 1.10 and 1.52 in the "full model". The strongest bivariate correlation was between age and eGFR (r = -0.49). Thus, no apparent multicolinearity was present and no further action was taken.

7. Interactions – additivity assumption

A two-way interaction occurs when the effect of one predictor is dependending on the value of one other predictor. There are several recommendations regarding the number of interactions to test for. Only clinically plausible interactions could be tested, however, this requires prior knowledge. Another strategy is to test for all possible interactions, this requires a very large sample, to avoid overfitting. A compromise is to do a pooled interaction test for each variable and if the test is significant, the specific interactions are pursued [11].

In the study
We did not have prespecified interactions for ADL and the sample size did not permit testing for all possible interactions. Therefore we opted for a global test approach. As we did not want to give ADL any advantages compared to the other variables, we also performed global tests for the other variables, one at a time. In the "model without ADL", the global test was significant for sex and BMI and an interaction term of sex * BMI was found (low BMI was a risk factor in men, not in women). This interaction was included in the model. In the "full model" another interaction, GBS-ADL*eGFR, was also found (the effect of impaired GBS-ADL was higher when eGFR was less imparied and vice versa). One interpretation of this interaction could be that impaired GBS-ADL is associated with weight loss and thus lower eGFR. To test properly for this we would need to apply three-way interactions (such as GBS-ADL*BMI*eGFR), which was beyond the scope of this paper.

8. Assumption of proportional hazards

The assumption of proportional hazards is the assumption that hazards from predictors do not vary over time. Proportional hazards can be tested in several different ways. Graphically, schoenfeld residuals are often plottet against time, then a straight line at zero is ideal. There are also different approaches to compensate for non-proportional hazards, the most common being adding an interaction term with time.

In the study

The proportional hazards assumption was first tested using a global test (`cox.zph` in R) as well as specific tests for all variables. In the "model without ADL", the global test gave a p value of 0.72 and in the "full model" a p value of 0.70, signalling no violations of the PH assumption. The variable closest was eGFR, with a p value of 0.14. For eGFR, a schoenfeld residual plot is shown in figure e5. No further action was taken.

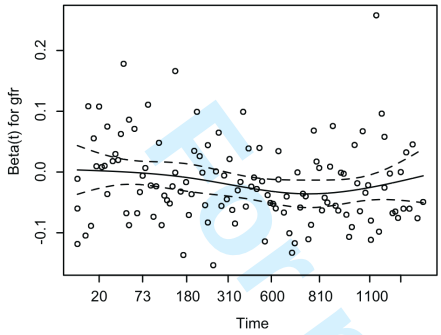

figure e5. Schoenfeld residual plot for eGFR.

9. Influential observations

With small sample size, a few influential observations could affect a model significantly. One way to screen for influential observations is by using what is called dfBeta, that shows to what extent the regression coefficient would change, if that case should be removed. Every case is designated a dfBeta value for each variable. Often, standardised dfBetas, with a cutoff of 0.20 is used to signify an influential observation. Thus, if deleting one observation led to a change in a predictor's β coefficient of more than 0.2 standard error, that observation was noted. For variables of specific interest, a sensitivity analysis could be performed without the observations with dfBeta > 0.2 to determine whether the effect is mainly due to a few highly influential observations.

In the study

In the "model without ADL", a total of 23 (12%) observations had any DfBeta > ±0.20. The lowest dfBeta was -0.39 and the highest 0.32. In the "full model", 21 observations were considered influential. DfBetas ranged from -0.46 to 0.48. Nine cases had a dfBeta > ± 0.20 for GBS-ADL and/or its interaction with eGFR. A sensitivity analysis was done, with these nine observations excluded. In that model the overall χ^2 increased from 123 to 124 and the GBS-ADL χ^2 from 32 to 37. Thus, the effects of GBS-ADL in the "full model" were not due to a few influential observations. In all further analysis the influential observations were kept in the model.

10. Relative contribution of variables

Describing the main effects of predictors including non-linear terms and interaction terms is not as intuitive as for simpler models, using Hazard Ratios. This is especially true if the model contains continuous-by-continuous interactions.

In the study

To obtain an estimate of the relative importance of the different predictors, we used the anova approach, developed by Harrell (`anova.rms` in R)[11]. Simple anova plots were included in

the article as figure 1. Plots of the variable effects are shown below in figure e6. In these plots, interaction terms have been incorporated into the variables' relative importance.

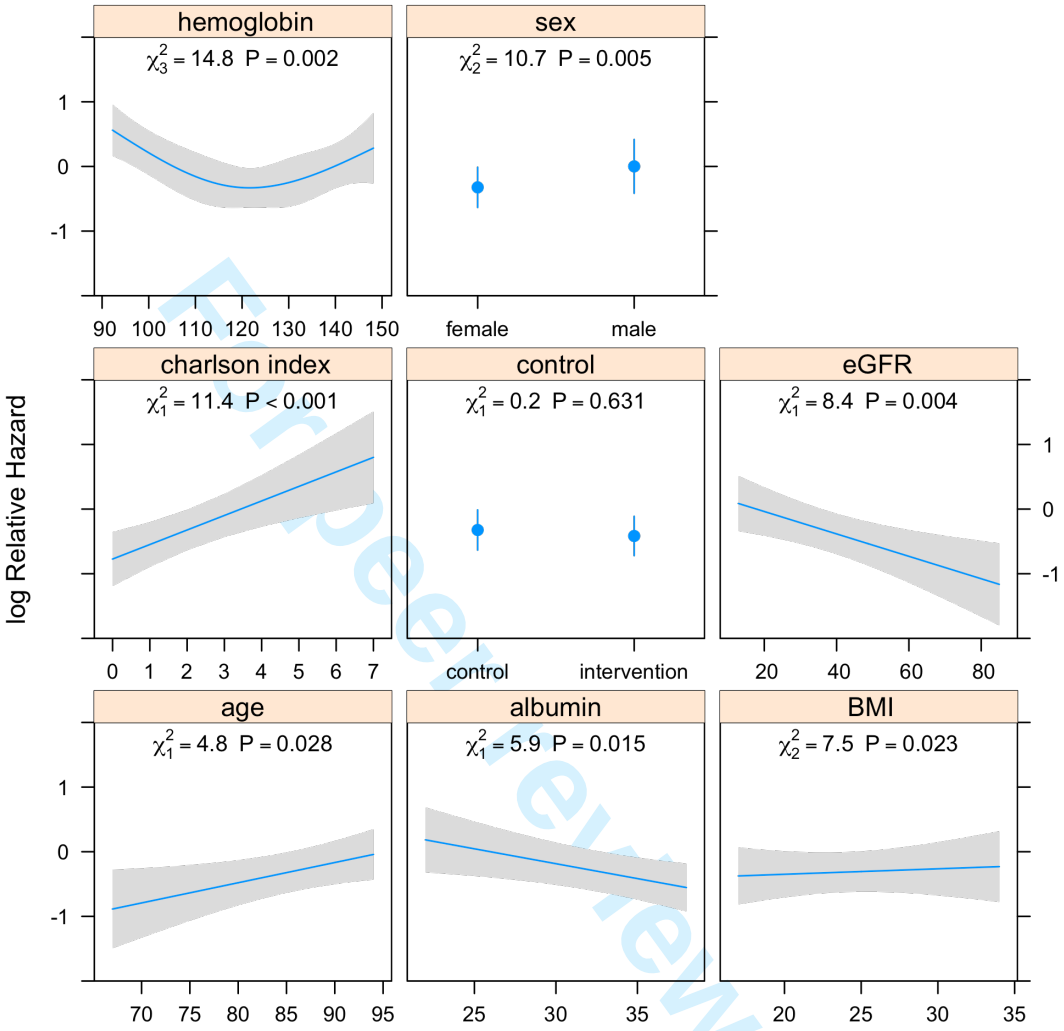

Figure e6a. Plot of variable effects in the "model without ADL"

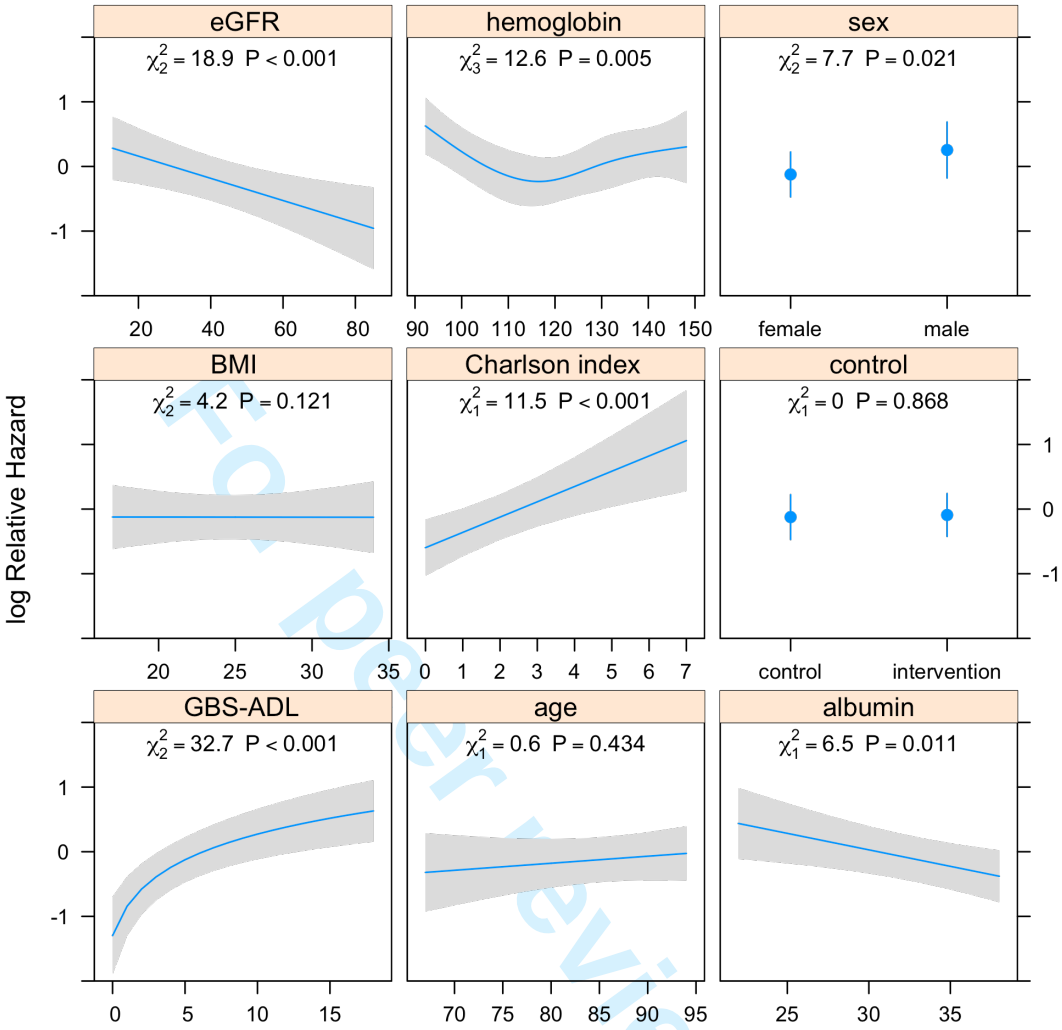

Figure e6b. Plots of the variable effects in the "full model".

11. Added value of an added variable

There are several ways to determine the added value of a variable in a regression model.

a. Likelihood ratio test. With two nested models (where the smaller model is also a part of the full model) a Likelihood ratio test could be performed as a χ^2 test over df = number of additional independent variables in the new model.

In the study

The results are shown in table 3 in the article. For the "model without ADL", LR χ^2 was 78.4 and for the "full model" 121.0. The degrees of freedom were 11 and 13, respectively. Therefore the LR test resulted in a χ^2 (df = 2, N = 198) = 42.5, p < 0.001. Thus the "full model" had a significantly better fit.

b. Discrimination, measured with the C, or concordance, statistic. The C statistic is the probability that, in a case-control pair, the case (deceased) will be given a higher predicted risk from the model than the control (survivor). C statistics range from 0.5 (coin toss, useless) to 1.0 (perfect discrimination). In logistic regression (without time-to-event data), the c

statistic is the same as ROC. For survival analysis, time is incorporated, so a case at time t is compared with a survivor at time t, albeit this survivor could be dead at time t+1 (the next day). C statistics in survival analysis are often lower than ROC in logistic analysis. In addition, there are several different ways to calculate c statistic for time-to-event data.

In the study
We chose the method by Uno, to be able to compare between models. The "model without ADL" had a c statistic of 0.72 and the "full model" of 0.78. We set the follow-up time to 1428 days, as this was our median follow-up time of survivors. C statistics from the two models were compared using the method described by Uno et al. in the SurvC1 package [13]. Difference in c statistic between the model without ADL and the full model was 0.058 (95% CI = 0.022 - 0.094, p value 0.002).

c. NRI >0. Continuous net reclassification index[14, 15]. This index determines to what extent adding a new variable to a model leads to a change in the correct direction in predicted risk for each observation at time t (towards higher risk for deceased, towards lower for survivors). NRI>0 ranges from 0 (no increased value, useless) to 1(all observations reclassified in the right direction). NRI>0 has been shown to be more sensitive than change in C index, especially when the baseline model has a good performance. NRI>0 only describes the share of observations that have been reclassified, it does not quantify the amount of change in risk. Thus, it cannot distinguish between adding a variable that increases the predicted mortality risk for all cases with 1% or one that increases it with 50%.

For interpretation, the original NRI > 0 has been compared to the effect size of the added variable, where NRI>0 of 0.6 should be considered strong, 0.4 intermediate and below 0.2 weak [16]. However, after the initial development, Pencina et al. have suggested that ½ NRI>0 shoud be reported, as an average[15]. This is also what is given by the `IDI.INF` function in the `SurvIDINRI` package in R.

In the study
In our study ½ NRI>0 (95%CI) was 0.42 (0.22-0.58) with a p value <0.001. Again the follow-up time was set to 1428 days, to avoid extensive censoring. By doubling the point estimate of ½ NRI>0, the original NRI>0 would be 0.84, indicating a substantial effect size of adding ADL.

IDI. Integrated discrimination improvement. Originally developed by Pencina et al. for logistic models, IDI has been extended to time-to-event data [14, 17]. While NRI>0 displays the percentage of observtations being reclassified in the desirable direction, IDI is related to mean change in predicted probabilities within cases and controls. IDI is similar to testing the difference in R2, or discrimination slope, in logistic regrssion. IDI and NRI with confidence intervals were calculated (using the `IDI.INF` function) with the method by Uno et al. [18].

In the study
IDI was 0.15 (95%CI 0.07-0.27, p < 0.001), indicating that the mean change in the correct direction was 15% (cases (deceased) were given 15% higher mortality risk by adding ADL while controls (survivors) were given 15% lower).

12. Overfitting - internal validation

Overfitting occurs when sample size is small. Then the fitted model becomes too optimistic and dependent on the present dataset. Thus, the findings will neither be reproducable nor

valid in other populations. Ideally a model could be tested in another population at another location and setting, what is known as external validation. If that is not possible, there are several ways to accomplish internal validation. The recommended approach is via bootstrapping. In bootstrapping a new dataset, of the same size as the original, is constructed from the original dataset by resampling with replacement (an observation could be selected several times). This dataset is then used to develop the model, which is then tested on the full original population. The difference in apparent performance and resampled performance is called optimism. The procedure is repeated 200-1000 times and the mean optimism is subtracted from the apparent performance estimates. This way an optimism-corrected estimate is obtained. In a future external population, this corrected estimate should be considered the best estimate possible[2, 8, 11, 19, 20].

In the study

Our aim was not to develop a valid prediction model, as this would have called for a larger sample size. Instead, we aimed to determine the relative imporance and added value of ADL when compared to the best possible traditional model. In the trade-off between overfitting and a well-performing traditional model, we empahsised the latter. A heuristic estimate of shrinkage would be χ^2 - d.f. / χ^2 = 123 - 14/123 = 0.89 for the "full model". However, this d.f. is falsely low as we tested many more interactions and transformations.

To better determine, the extent of overfitting, we carried out an internal validation with 1000 bootstraps. For the "model without ADL" and the "full model", the calibration slopes were 0.84 and 0.83, respectively, indicating a substantial amount of overfitting. The optimism-corrected R^2 was 0.26 vs. 0.39. Optimism-corrected c statistics were 0.69 and 0.76.

13. Updating the model - final nomogram and Kaplan-Meier curves

An overfitted model could be updated using a model that shrinks the regression coefficients. One such method is the LASSO (least absolute shrinkage and selection operator) model [21, 22]. LASSO could be used to both shrink factors as well as to eliminate variables.

In the study

Even if our aim was not to develop a valid prediction model, the amount of overfitting suggested that we should try to update the "full model". In a LASSO model, the interaction terms and non-linear terms were combined into single terms. We used the `coxpath` function in R. We considered the model with the lowest AIC (Akaike Information Criterion). In this model, the variable "control" was shrunk to zero, all other variables remained. The mean shrinkage was 0.84. The lasso path and AIC is shown in figure e7.

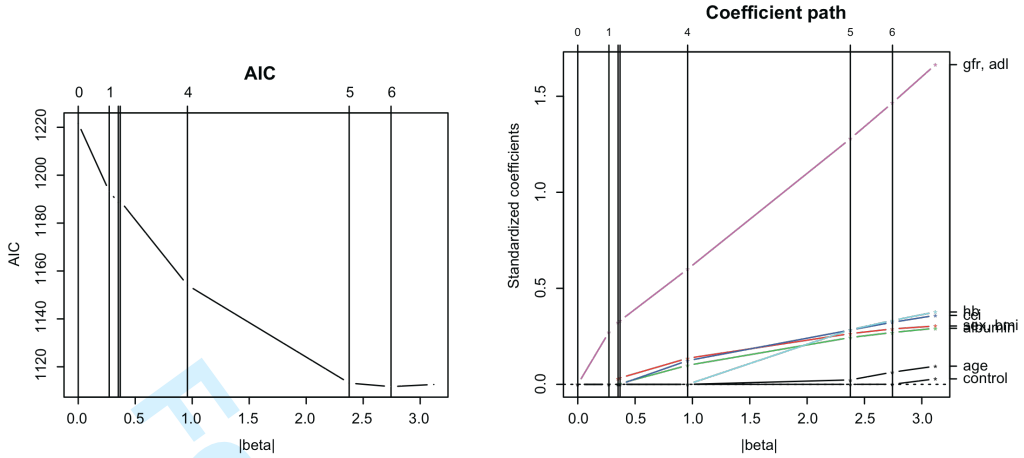

Figure e7. Lasso plots. AIC (Akaike information criterion) is lowest when control is set to zero.

A final nomogram, using the shrunk Lasso coefficients, was built, see figure e8. The cases were divided into four equally sized risk groups, by the quartiles of the linear predictor. To display the discrimination of the model, a kaplan-Meier curve was built, included in the article as figure 2.

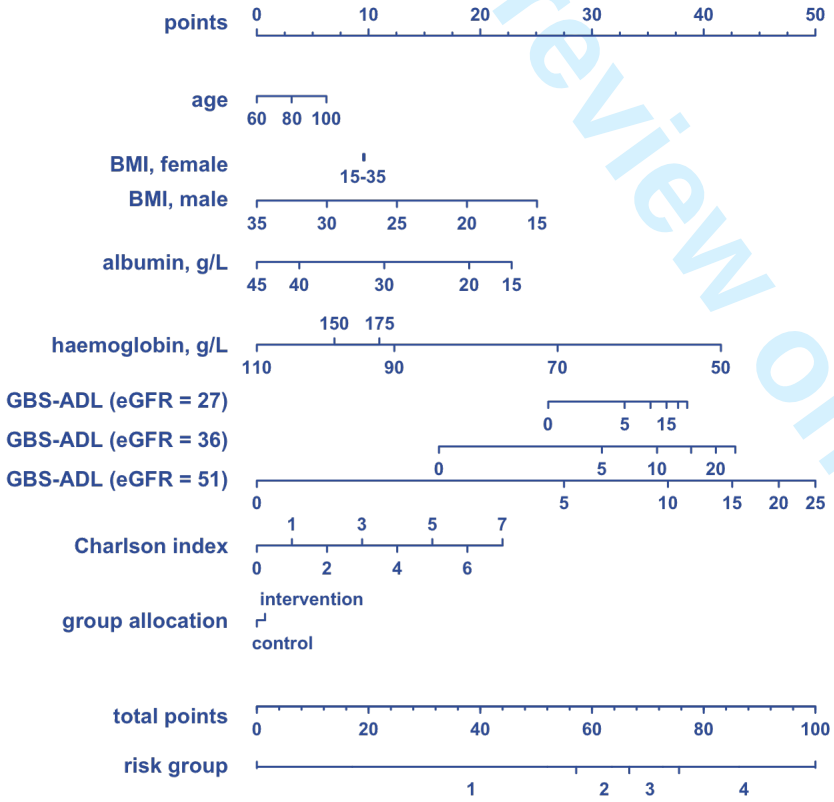

Figure e8a. Nomogram. Interpretation: For an individual, the variables are compared with the upper ”points” line, one at a time. These scores are then added for a total score that is plotted at the ”total points” line at the bottom. This could then be used to designate the person to a ”risk group” Notice the effect of interactions, low BMI is only a risk factor in men and the risk of GBS-ADL is moderated by eGFR, which is presented by median and quartiles. The cutoffs

in the nomogram for the risk groups are completely arbitrary here, created to obtain 4 equally sized groups. In another scenario, cutoffs could be established to obtain for example a group with 90% chance of 3-year survival.

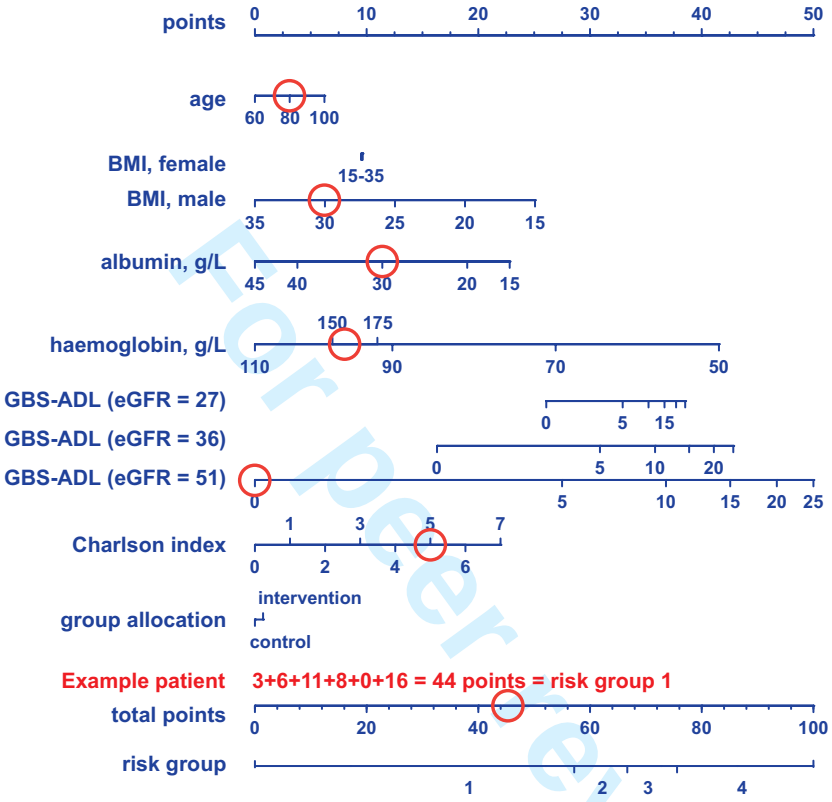

Figure e8b. Example of a scoring: This patient is 80 years old (3 points), male with BMI 30 (6 points), has an albumin of 30 (11 points) and a hemoglobin of 98 (8 points), normal kidney function and GBS-ADL (0 points) and a Charlson index score of 5 (16 points). The total score would be 3+6+11+8+0+16 = 44 points, placing this patient well within risk group 1. If this patient had all other variables constant but a functional decline, with a GBS-ADL score of 7, this would result in a total score of 44+30 = 74, placing the patient in risk group 3. The risk attributed to the functional decline would be equivalent to a hemoglobin drop from 98 to 55 g/L. Would it infer the same sense of urgency to the clinician?

References - statistical appendix

1. *R: A language and environment for statistical computing.* [computer program]. Version. VIenna; 2014.
2. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* Feb 28 1996;15(4):361-387.
3. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol.* Oct 2006;59(10):1087-1091.
4. Horton NJ, Kleinman KP. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *Am Stat.* Feb 2007;61(1):79-90.
5. Janssen KJ, Donders AR, Harrell FE, Jr., et al. Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol.* Jul 2010;63(7):721-727.
6. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Med Res Methodol.* 2010;10:7.
7. Moons KG, Donders RA, Stijnen T, Harrell FE, Jr. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol.* Oct 2006;59(10):1092-1101.
8. Steyerberg EW. *Clinical Prediction Models - A Practical Approach to Development, Validation, and Updating*. New York: Springer; 2009.
9. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med.* Jan 15 2006;25(1):127-141.
10. Harrell FE, Jr., Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst.* Oct 5 1988;80(15):1198-1202.
11. Harrell Jr FE. *Regression Modeling Strategies - With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer-Verlag; 2001.
12. O'Brien RM. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity* 2007-10-01 2007;41(5):673-690.
13. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med.* May 10 2011;30(10):1105-1117.
14. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med.* Jan 30 2008;27(2):157-172; discussion 207-112.
15. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med.* Jan 15 2011;30(1):11-21.
16. Pencina MJ, D'Agostino RB, Pencina KM, Janssens AC, Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol.* Sep 15 2012;176(6):473-481.

17. Chambless LE, Cummiskey CP, Cui G. Several methods to assess improvement in risk prediction models: extension to survival analysis. *Stat Med.* Jan 15 2011;30(1):22-38.

18. Uno H, Tian L, Cai T, Kohane IS, Wei LJ. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Stat Med.* Jun 30 2013;32(14):2430-2442.

19. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med.* Aug 30 2004;23(16):2567-2586.

20. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* Aug 2001;54(8):774-781.

21. Tibshirani R. Regression shrinkage and selection via the Lasso. *J. Royal. Statist. Soc B.* 1996;58(1):267-288.

22. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med.* Feb 28 1997;16(4):385-395.

Appendix 1 –statistics

General aspects

The overall aim was to compare ADL with the best possible model containing customary risk factors.
To achieve this, two models were developed and compared, the "model without ADL" and the "full model".
Data was originally stored in an SPSS file. All data analysis was performed in R [1]. Code is provided in appendix 2.

1. Outcome

Generally, it is important to describe the quantity of cases with missing outcome and to determine if there are any underlying patterns. Otherwise, simple exclusion may affect representativity [2].

In the study

Survival status was determined using the local region's electronic registry on the 6th February 2014. The time variable was defined as days from discharge to death or censoring at study endpoint, whichever came first. Those surviving at endpoint had been followed for a median of 1428 days (range 1312-1548). The baseline survival function is shown in figure e1.

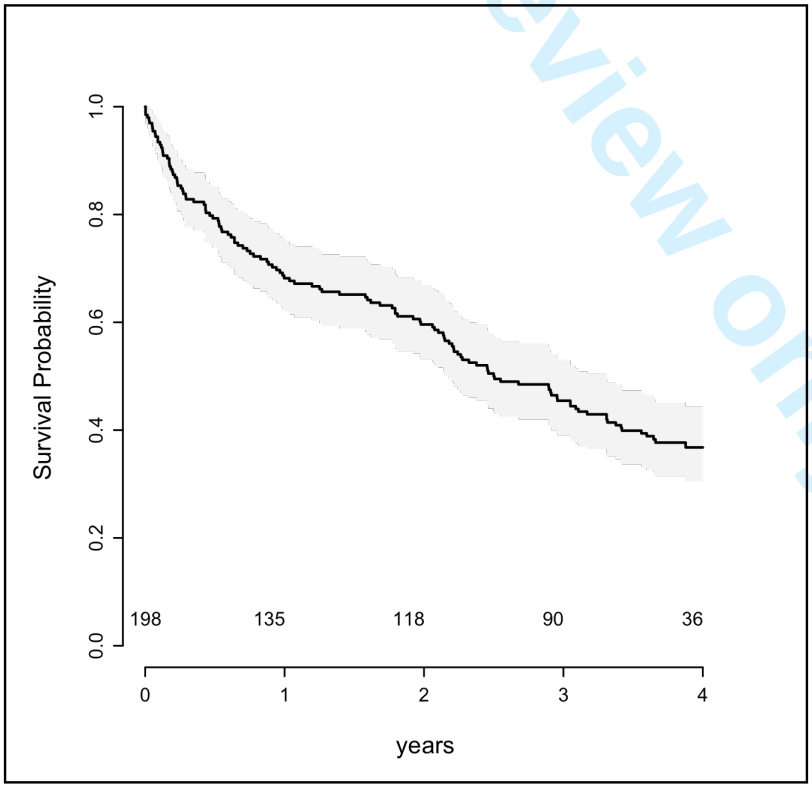

figure e1. Baseline survival function.

In our study, two cases were missing survival status due to having moved abroad (no longer in the region's registry) Hypothetically, these cases could be assumed to be in better health (severely diseased patients are unlikely to move abroad?). However, they were considered too few to affect representativity and were discarded from further analysis. Thus, the number of cases decreased from 200 to 198.

2. Crude analysis

Before any modifications are done to a variable, a crude analysis for the intented outcome could be of interest, to obtain an initial estimate of the effect of the predictor

In the study

Bivariate Cox proportional hazards regressions were carried out separately for all variables, including only outcome and the variable. All variables were treated in their original form, on their original scale. Observations with missing values were excluded from crude analysis. Data is presented with β coefficients, Standard errors, Wald χ^2, p value and hazard ratios in table 2 in the article.

In the crude analysis, all variables/potential predictors were statistically significant except sex, control/intervention status in the original study and the ADL item "food intake". Regarding the latter, the distribution was severely skewed, with only 18 cases (9%) having a non-zero value. To obtain a preliminary ranking of importance, the variables were sorted by decreasing Wald χ^2 in the table in the article.

In crude analysis, all separate GBS-ADL items were included but in further multivariate analysis, only the total GBS-ADL score was used, to avoid fitting too many variables and multicolinearity (the ADL items were intercorrelated at r = 0.8-0.9)

3. Missing data

In general, it is important to analyse missing data patterns in predictors. The first step is to determine the quantity of missing data. The second step is whether data is missing completely at random or if there are underlying patterns. When these prerequisites have been fulfilled, there are several approaches to missing data:

1. Listwise deletion, discarding all observations with any missing data points. The advantage of this approach is that no "manipulation" is done. Therefore, this method may seem intuitively most correct. The obvious disadvantage is that sample size could be substantially diminished. In addition, representativity could be affected, if missing a variable is systematically associated with other characteristics.

2. Using simple imputation. This technique substitutes missing values with the mean, mode or median value. This could be acceptable only if the variable is missing completely at random and the percentage of missing values small.

3. Using a more complex imputational technique. This approach uses customised regression models including all other covariates to obtain a stable prediction of the missing values. This method has been described and emphasized in several publications [3-7]. When using complex imputations, single or multiple imputations could be chosen. In the latter case, a separate

dataset is analysed for each imputational iteration, leading to a much larger complexity in the analysis.

In the study

When analysing the quantity of missing data, eGFR was missing in one case, BMI in three cases. Albumin was missing in 17 (9%) cases. BNP was missing in 113 (56%) observations.

The BNP variable was discarded from further analysis, as it had more than 50% missing. BMI and eGFR were considered missing completely at random. However, we found that cases with missing albumin were predominantly female (15 female vs. 2 male, $\chi^2 = 3.32$, $p = 0.056$) and had lower score on Charlson comorbidity index (1.47 vs. 2.33, $F = 11.3$, $p = 0.002$). Thus, excluding cases with missing albumin would affect representativity. Discarding the albumin variable would affect the overall aim, to compare ADL with the best possible traditional model. Therefore, the missing values in BMI, eGFR and Albumin were imputed using a single conditional imputation method (with the `transcan` function in R). In total, the effect of imputations was very small on the variable properties, as shown below.

Variable	β	S.E	Wald X2	p value	HR (95% CI)
Albumin, g/L, n = 181	-0.064	0.018	13.1	<0.001	0.94 (0.90 - 0.97)
- with imputation, transcan	-0.066	0.017	14.7	<0.001	0.94 (0.91 - 0.97)
eGFR, ml/min, n = 197	-0.029	0.005	29.3	<0.001	0.97 (0.96 - 0.98)
- with imputation, transcan	-0.029	0.005	29.2	<0.001	0.97 (0.96 - 0.98)
BMI, kg/m^2 , n = 195	-0.053	0.020	7.4	0.007	0.95 (0.91 - 0.99)
- with imputation, transcan	-0.053	0.020	7.4	0.006	0.95 (0.91 - 0.99)

Table e1. Effect of imputation on variable properties.

4. Variable considerations

Extreme outliers

In regression, outliers may be defined as observations with more than 3 interquartile ranges over the third quartile or below the 1st quartile. Such extreme values may affect a regression model significantly. First data entry errors should be considered and pursued. Then the biological plausibility should be considered. If plausible, we may consider a truncation at the 99th or 1st percentile [8].

In the study

In our study, data screening revealed, that for eGFR there was one extreme outlier with an estimated value of 198 ml /min (> 6 IQR over 3rd quartile), see boxplot.

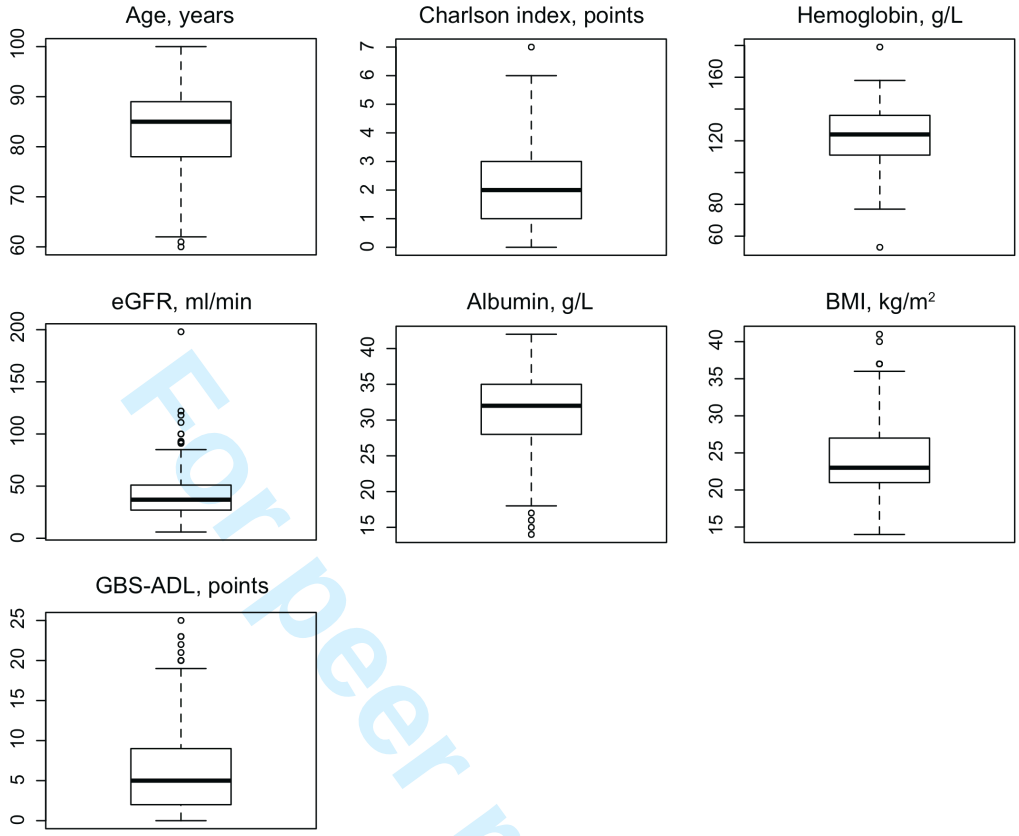

Figure e2. Boxplots of the continuous predictors. eGFR = Glomerular filtration rate, BMI = Body mass index, ADL = Activities of daily living.

This case was screened for data entry errors but none were found. Regarding biological plausibility, eGFR was measured with the Cockcroft-Gault formula ((140-age) * weight * constant)/Serum Creatinine in μmol/L, where the constant is 1.23 for men and 1.04 for women. Thus, GFR was not measured directly, but estimated and sensitive to extreme values in both serum creatinine, age and body weight. With this reservation, we considered the value to be biologically plausible. However, we did not consider it clinically important to compare one elderly patient with 198 ml/min in eGFR with another with 120 ml/min with regard to mortality. Therefore, eGFR was winsorized at the 99th percentile (118 ml/min). This led to a slightly improved fit in univariate performance.

Variable	β	S.E	Wald X2	p value	HR (95% CI)
eGFR, ml/min	-0.029	0.005	29.2	<0.001	0.97 (0.96 - 0.98)
- winsorized at 99th percentile	-0.029	0.005	29.8	<0.001	0.97 (0.96 - 0.98)

Table e2. effect of winsorization on variable properties.

Non-linearity

Most regression model assume that the predictors are linearly related to the outcome. However, non-linear relationships, such as U-shapes, for continuous variables are common.

There are several ways to address non-linearity:

First, assuming that the variable is linear. The advantage of this approach is that it results in an easily interpreted main effect, for example the Hazard Ratio in survival analysis. This is

the approach used in our crude comparisons. However, the approach is potentially problematic. For hemoglobin, this would mean that the risk difference between two individuals with 170 and 130 g/L would be the same as between two with 90 and 50, respectively. In addition, this approach cannot handle U-shaped risks, it is likely that someone with 200 g/L in Haemoglobin with dehydration or polycytemia does not have better survival than someone with 140 g/L

Second, to dichotomize the variable, using a previously established cut-off, is another frequently used approach. However it is not recommended as it ignores a lot of information [9]. In our example, applying the WHO cut-off for anemia (120 g/L for women, 130 for men) would attribute the same risk for an individual with Hb of 119 g/L as for one with 53 g/L (the lowest in our material).

Third, to categorise the variable into categories that are clinically important, creating dummy variables. This approach could handle U-shaped risks. However, previously defined clinically important categories are needed and several degrees of freedom is spent in the analysis. As with a dichotomous transformation, all cases within a category are attributed the same risk.

Fourth, to use a more complex fitting function, such as a restricted cubic spline [10, 11]. This approach uses so called knots, point estimates where the risk is determined. A cubic function is used to fit the function between knots. Near the ends the risk is modelled linear.

In the study

We prespecified Hemoglobin to be non-linear and tried the approaches above, see figure e4. We decided to use the 4-knot restricted cubic spline as both the best performance and was most appropriate from a clinical perspective. The knots were placed at the 5th, 35th, 65th and 95th percentiles where Hb was 92.25, 118, 130 and 148.15, respectively. The resulting function to fit Hb was:

```
2.2712251-0.017758194* hb-6.2295666e-06*pmax(hb-
92.25,0)^3+5.8240197e-05*pmax(hb-118,0)^3-7.7559735e-
05*pmax(hb-130,0)^3+2.5549104e-05*pmax(hb-148.15,0)^3
```

As opposed to the easily interpreted hazard ratio from the linear function, this is not easy to interpret without a graph, the graphic display of the four approaches is presented in figure e3.

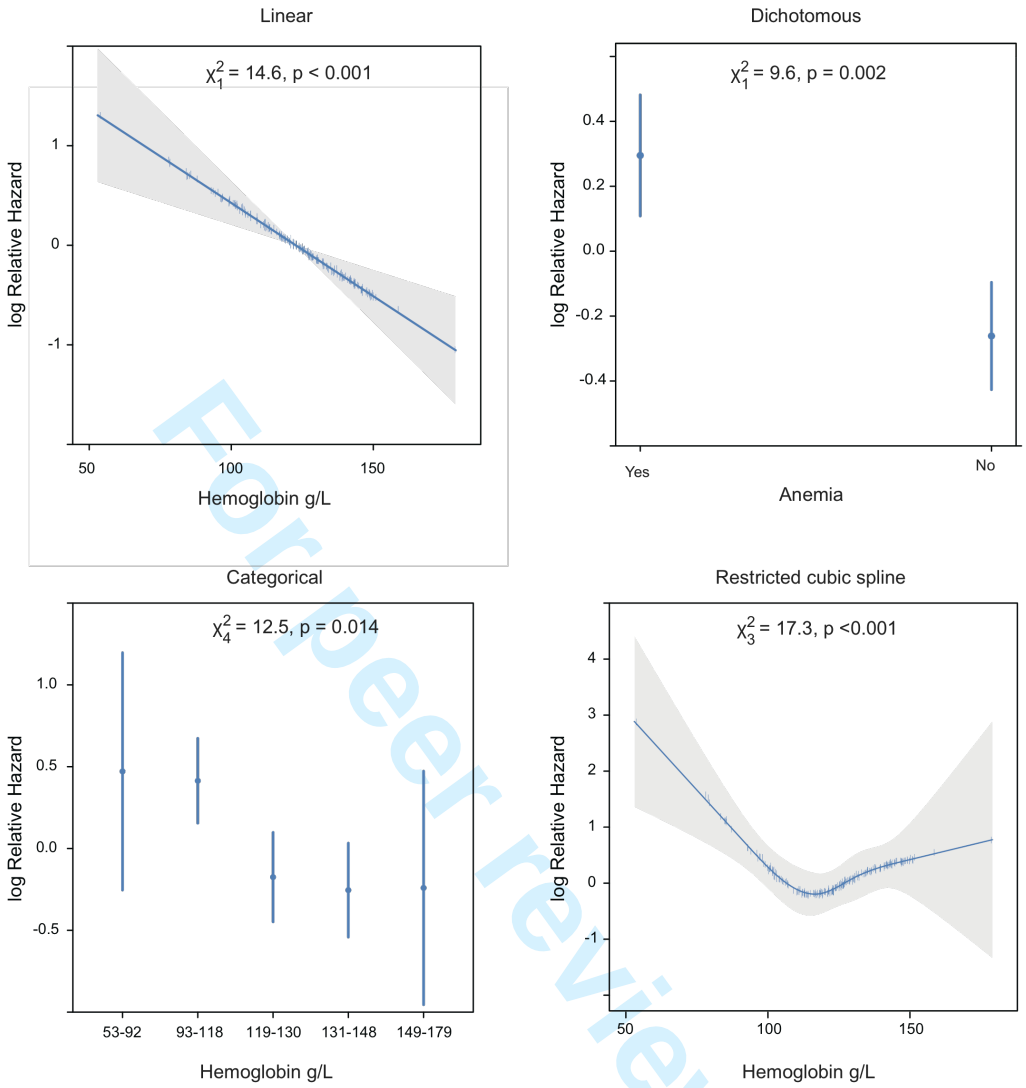

Figure e3. Different transformations of Hemoglobin. For dichotomous, the WHO definition of anemia is used. For categorical, the 5th, 35th, 65 and 95th percentiles were used, for easier comparison with the spline fit.

Apart from Hemoglogin, all other variables were bivariately tested for non-linearity by using 4-knot splines followed by ANOVA tests to determine if there was a significant non-linear component. GBS-ADL showed significant non-linearity and different codings were tested. We tested dichotomizing at the median and categorizing at the quartiles. A polynomial showed good fit but was not clinically plausible, with decreasing risks at the higher end of ADL impairment. The restricted cubic spline resembled a log fit and indeed the log fit was chosen, with fewer degrees of freedom spent, see figure e4. No other variables showed significant non-linear effects.

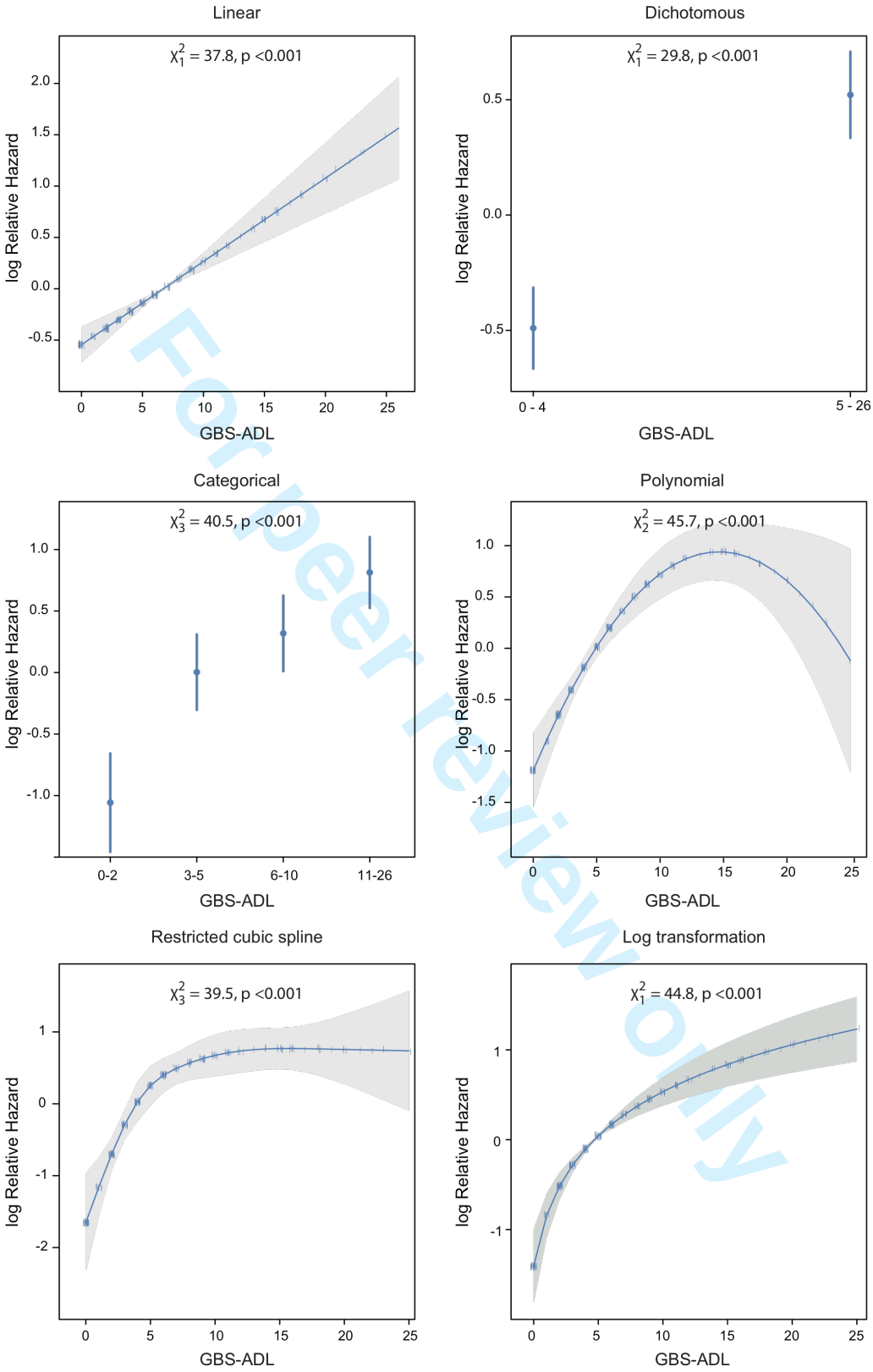

Figure e4. Different transformations tested for GBS-ADL.

5. Fitting the multivariate models

In the study

The two models were fitted, using the imputations and transformations above. The "model without ADL" used the covariates age, sex, charlson comorbidity index, albumin, BMI, eGFR, control/intervention status, and hemoglobin fitted as a restricted cubic spline The "full model" also included log(GBS-ADL).

6. Multivariate Diagnostics - Multicolinearity

Predictors with strong intercorrelations could cause interpreting problems, this is tested using the variance inflation factor (VIF). The interpretation of VIF has been disputed, a rule of thumb saying that VIF > 4 or > 10 signals a problematic multicolinearity problem have been suggested. However, these cut-offs may be too low, as a VIF over 10 could be acceptable [12]. To address multicolinearity, clustering of variables or data reduction could be applied.

In the study

In our models, all variables were simultaneously tested for colinearity. VIF Values were ranging between 1.02 and 1.47 in the "model without ADL" and between 1.10 and 1.52 in the "full model". The strongest bivariate correlation was between age and eGFR (r = -0.49). Thus, no apparent multicolinearity was present and no further action was taken.

7. Interactions – additivity assumption

A two-way interaction occurs when the effect of one predictor is dependending on the value of one other predictor. There are several recommendations regarding the number of interactions to test for. Only clinically plausible interactions could be tested, however, this requires prior knowledge. Another strategy is to test for all possible interactions, this requires a very large sample, to avoid overfitting. A compromise is to do a pooled interaction test for each variable and if the test is significant, the specific interactions are pursued [11].

In the study

We did not have prespecified interactions for ADL and the sample size did not permit testing for all possible interactions. Therefore we opted for a global test approach. As we did not want to give ADL any advantages compared to the other variables, we also performed global tests for the other variables, one at a time. In the "model without ADL", the global test was significant for sex and BMI and an interaction term of sex * BMI was found (low BMI was a risk factor in men, not in women). This interaction was included in the model. In the "full model" another interaction, GBS-ADL*eGFR, was also found (the effect of impaired GBS-ADL was higher when eGFR was less imparied and vice versa). One interpretation of this interaction could be that impaired GBS-ADL is associated with weight loss and thus lower eGFR. To test properly for this we would need to apply three-way interactions (such as GBS-ADL*BMI*eGFR), which was beyond the scope of this paper.

8. Assumption of proportional hazards

The assumption of proportional hazards is the assumption that hazards from predictors do not vary over time. Proportional hazards can be tested in several different ways. Graphically, schoenfeld residuals are often plottet against time, then a straight line at zero is ideal. There are also different approaches to compensate for non-proportional hazards, the most common being adding an interaction term with time.

In the study

The proportional hazards assumption was first tested using a global test (`cox.zph` in R) as well as specific tests for all variables. In the "model without ADL", the global test gave a p value of 0.72 and in the "full model" a p value of 0.70, signalling no violations of the PH assumption. The variable closest was eGFR, with a p value of 0.14. For eGFR, a schoenfeld residual plot is shown in figure e5. No further action was taken.

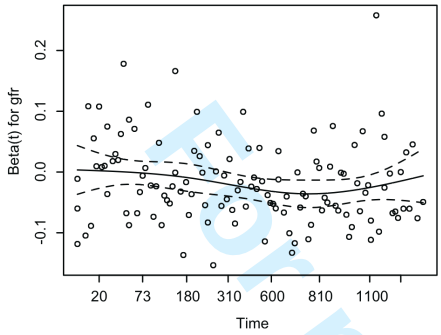

figure e5. Schoenfeld residual plot for eGFR.

9. Influential observations

With small sample size, a few influential observations could affect a model significantly. One way to screen for influential observations is by using what is called dfBeta, that shows to what extent the regression coefficient would change, if that case should be removed. Every case is designated a dfBeta value for each variable. Often, standardised dfBetas, with a cutoff of 0.20 is used to signify an influential observation. Thus, if deleting one observation led to a change in a predictor's β coefficient of more than 0.2 standard error, that observation was noted. For variables of specific interest, a sensitivity analysis could be performed without the observations with dfBeta > 0.2 to determine whether the effect is mainly due to a few highly influential observations.

In the study

In the "model without ADL", a total of 23 (12%) observations had any DfBeta > ±0.20. The lowest dfBeta was -0.39 and the highest 0.32. In the "full model", 21 observations were considered influential. DfBetas ranged from -0.46 to 0.48. Nine cases had a dfBeta > ± 0.20 for GBS-ADL and/or its interaction with eGFR. A sensitivity analysis was done, with these nine observations excluded. In that model the overall χ^2 increased from 123 to 124 and the GBS-ADL χ^2 from 32 to 37. Thus, the effects of GBS-ADL in the "full model" were not due to a few influential observations. In all further analysis the influential observations were kept in the model.

10. Relative contribution of variables

Describing the main effects of predictors including non-linear terms and interaction terms is not as intuitive as for simpler models, using Hazard Ratios. This is especially true if the model contains continuous-by-continuous interactions.

In the study

To obtain an estimate of the relative importance of the different predictors, we used the anova approach, developed by Harrell (`anova.rms` in R)[11]. Simple anova plots were included in

the article as figure 1. Plots of the variable effects are shown below in figure e6. In these plots, interaction terms have been incorporated into the variables' relative importance.

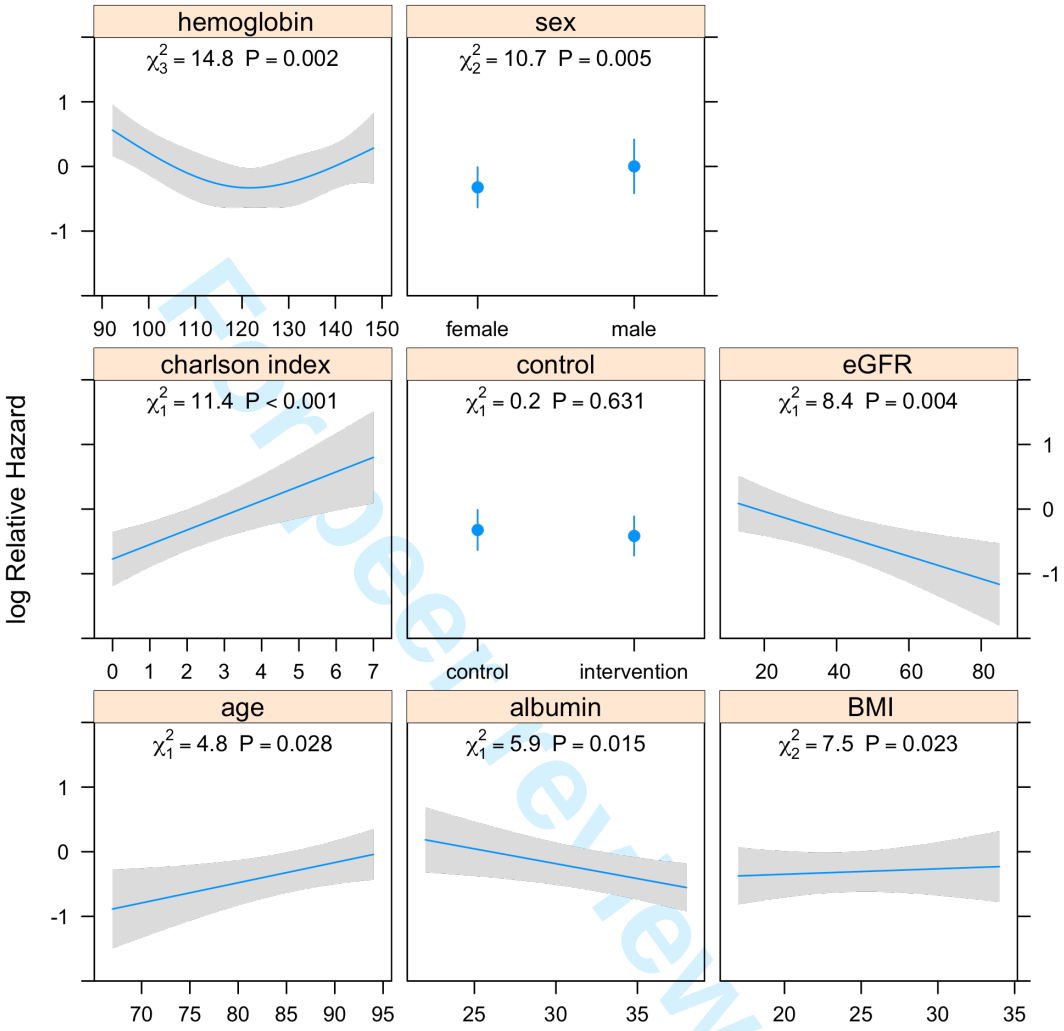

Figure e6a. Plot of variable effects in the "model without ADL"

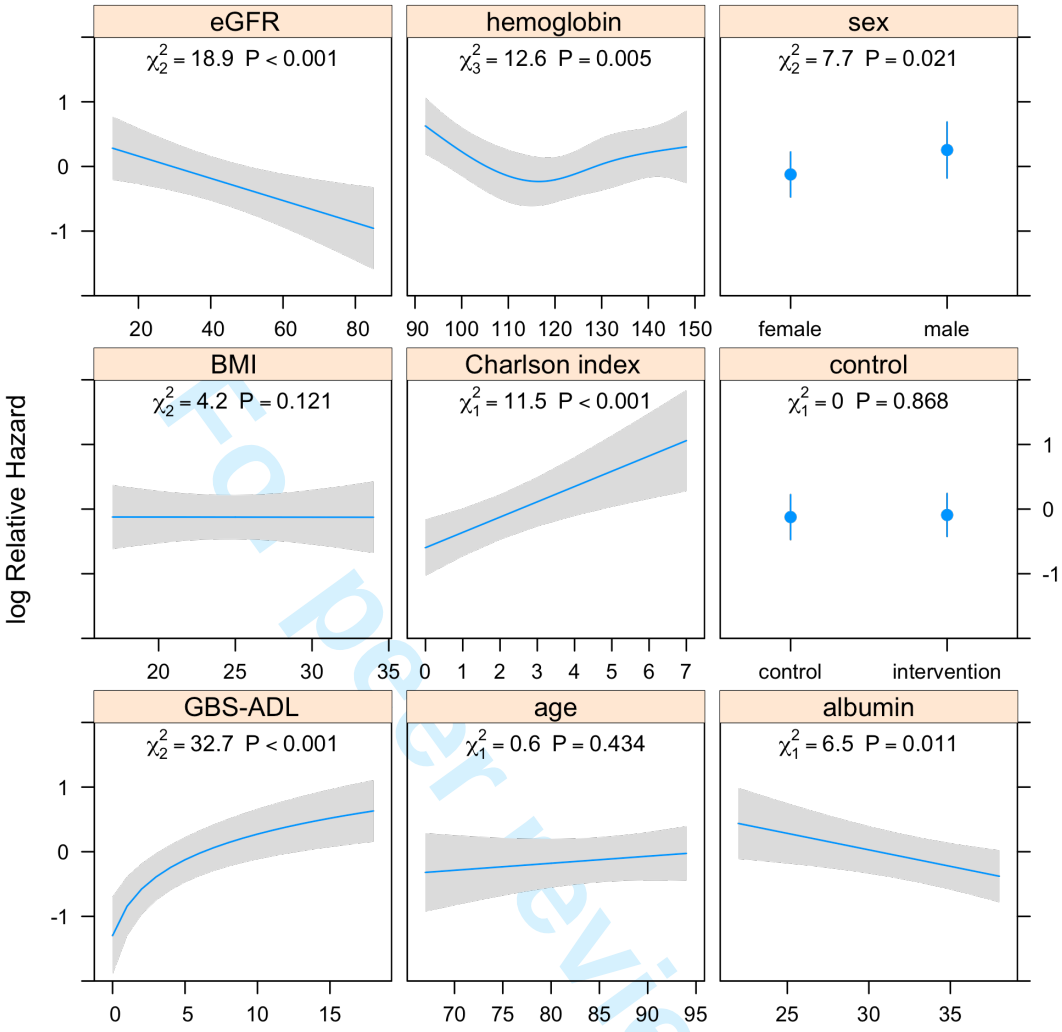

Figure e6b. Plots of the variable effects in the "full model".

11. Added value of an added variable

There are several ways to determine the added value of a variable in a regression model.

a. Likelihood ratio test. With two nested models (where the smaller model is also a part of the full model) a Likelihood ratio test could be performed as a χ^2 test over df = number of additional independent variables in the new model.

In the study

The results are shown in table 3 in the article. For the "model without ADL", LR χ^2 was 78.4 and for the "full model" 121.0. The degrees of freedom were 11 and 13, respectively. Therefore the LR test resulted in a χ^2 (df = 2, N = 198) = 42.5, p < 0.001. Thus the "full model" had a significantly better fit.

b. Discrimination, measured with the C, or concordance, statistic. The C statistic is the probability that, in a case-control pair, the case (deceased) will be given a higher predicted risk from the model than the control (survivor). C statistics range from 0.5 (coin toss, useless) to 1.0 (perfect discrimination). In logistic regression (without time-to-event data), the c

statistic is the same as ROC. For survival analysis, time is incorporated, so a case at time t is compared with a survivor at time t, albeit this survivor could be dead at time t+1 (the next day). C statistics in survival analysis are often lower than ROC in logistic analysis. In addition, there are several different ways to calculate c statistic for time-to-event data.

In the study

We chose the method by Uno, to be able to compare between models. The "model without ADL" had a c statistic of 0.72 and the "full model" of 0.78. We set the follow-up time to 1428 days, as this was our median follow-up time of survivors. C statistics from the two models were compared using the method described by Uno et al. in the SurvC1 package [13]. Difference in c statistic between the model without ADL and the full model was 0.058 (95% CI = 0.022 - 0.094, p value 0.002).

c. NRI >0. Continuous net reclassification index[14, 15]. This index determines to what extent adding a new variable to a model leads to a change in the correct direction in predicted risk for each observation at time t (towards higher risk for deceased, towards lower for survivors). NRI>0 ranges from 0 (no increased value, useless) to 1(all observations reclassified in the right direction). NRI>0 has been shown to be more sensitive than change in C index, especially when the baseline model has a good performance. NRI>0 only describes the share of observations that have been reclassified, it does not quantify the amount of change in risk. Thus, it cannot distinguish between adding a variable that increases the predicted mortality risk for all cases with 1% or one that increases it with 50%.

For interpretation, the original NRI > 0 has been compared to the effect size of the added variable, where NRI>0 of 0.6 should be considered strong, 0.4 intermediate and below 0.2 weak [16]. However, after the initial development, Pencina et al. have suggested that ½ NRI>0 shoud be reported, as an average[15]. This is also what is given by the `IDI.INF` function in the `SurvIDINRI` package in R.

In the study

In our study ½ NRI>0 (95%CI) was 0.42 (0.22-0.58) with a p value <0.001. Again the follow-up time was set to 1428 days, to avoid extensive censoring. By doubling the point estimate of ½ NRI>0, the original NRI>0 would be 0.84, indicating a substantial effect size of adding ADL.

IDI. Integrated discrimination improvement. Originally developed by Pencina et al. for logistic models, IDI has been extended to time-to-event data [14, 17]. While NRI>0 displays the percentage of observtations being reclassified in the desirable direction, IDI is related to mean change in predicted probabilities within cases and controls. IDI is similar to testing the difference in R2, or discrimination slope, in logistic regrssion. IDI and NRI with confidence intervals were calculated (using the `IDI.INF` function) with the method by Uno et al. [18].

In the study

IDI was 0.15 (95%CI 0.07-0.27, p < 0.001), indicating that the mean change in the correct direction was 15% (cases (deceased) were given 15% higher mortality risk by adding ADL while controls (survivors) were given 15% lower).

12. Overfitting - internal validation

Overfitting occurs when sample size is small. Then the fitted model becomes too optimistic and dependent on the present dataset. Thus, the findings will neither be reproducable nor

valid in other populations. Ideally a model could be tested in another population at another location and setting, what is known as external validation. If that is not possible, there are several ways to accomplish internal validation. The recommended approach is via bootstrapping. In bootstrapping a new dataset, of the same size as the original, is constructed from the original dataset by resampling with replacement (an observation could be selected several times). This dataset is then used to develop the model, which is then tested on the full original population. The difference in apparent performance and resampled performance is called optimism. The procedure is repeated 200-1000 times and the mean optimism is subtracted from the apparent performance estimates. This way an optimism-corrected estimate is obtained. In a future external population, this corrected estimate should be considered the best estimate possible[2, 8, 11, 19, 20].

In the study

Our aim was not to develop a valid prediction model, as this would have called for a larger sample size. Instead, we aimed to determine the relative imporance and added value of ADL when compared to the best possible traditional model. In the trade-off between overfitting and a well-performing traditional model, we empahsised the latter. A heuristic estimate of shrinkage would be χ^2 - d.f. / χ^2 = 123 - 14/123 = 0.89 for the "full model". However, this d.f. is falsely low as we tested many more interactions and transformations.

To better determine, the extent of overfitting, we carried out an internal validation with 1000 bootstraps. For the "model without ADL" and the "full model", the calibration slopes were 0.84 and 0.83, respectively, indicating a substantial amount of overfitting. The optimism-corrected R^2 was 0.26 vs. 0.39. Optimism-corrected c statistics were 0.69 and 0.76.

13. Updating the model - final nomogram and Kaplan-Meier curves

An overfitted model could be updated using a model that shrinks the regression coefficients. One such method is the LASSO (least absolute shrinkage and selection operator) model [21, 22]. LASSO could be used to both shrink factors as well as to eliminate variables.

In the study

Even if our aim was not to develop a valid prediction model, the amount of overfitting suggested that we should try to update the "full model". In a LASSO model, the interaction terms and non-linear terms were combined into single terms. We used the `coxpath` function in R. We considered the model with the lowest AIC (Akaike Information Criterion). In this model, the variable "control" was shrunk to zero, all other variables remained. The mean shrinkage was 0.84. The lasso path and AIC is shown in figure e7.

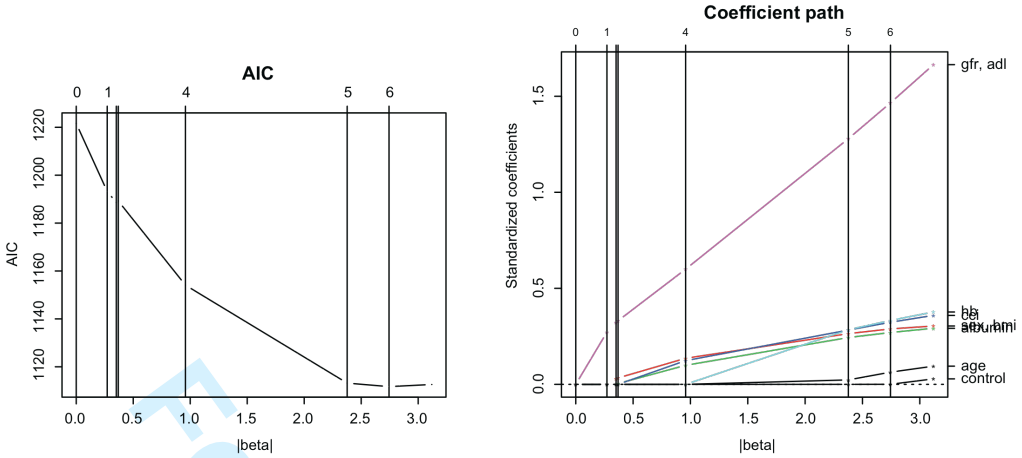

Figure e7. Lasso plots. AIC (Akaike information criterion) is lowest when control is set to zero.

A final nomogram, using the shrunk Lasso coefficients, was built, see figure e8. The cases were divided into four equally sized risk groups, by the quartiles of the linear predictor. To display the discrimination of the model, a kaplan-Meier curve was built, included in the article as figure 2.

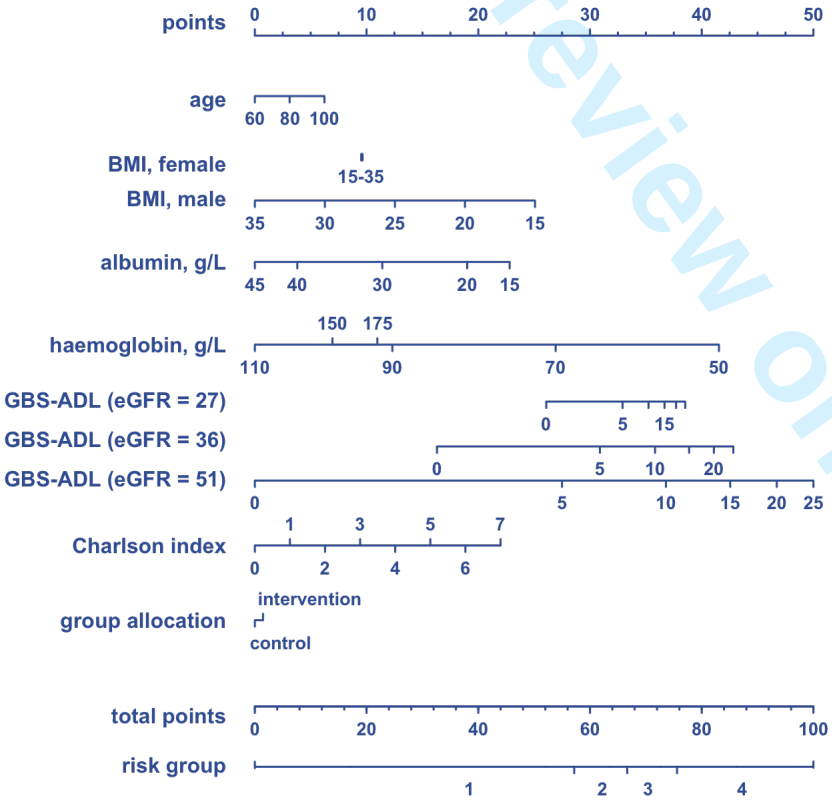

Figure e8a. Nomogram. Interpretation: For an individual, the variables are compared with the upper ”points” line, one at a time. These scores are then added for a total score that is plotted at the ”total points” line at the bottom. This could then be used to designate the person to a ”risk group” Notice the effect of interactions, low BMI is only a risk factor in men and the risk of GBS-ADL is moderated by eGFR, which is presented by median and quartiles. The cutoffs

in the nomogram for the risk groups are completely arbitrary here, created to obtain 4 equally sized groups. In another scenario, cutoffs could be established to obtain for example a group with 90% chance of 3-year survival.

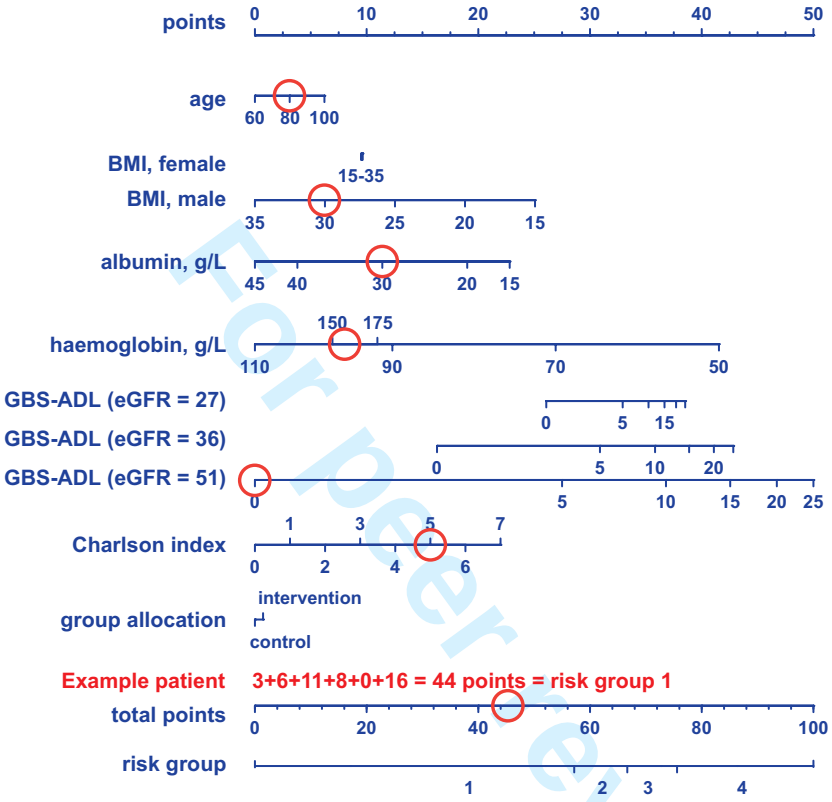

Figure e8b. Example of a scoring: This patient is 80 years old (3 points), male with BMI 30 (6 points), has an albumin of 30 (11 points) and a hemoglobin of 98 (8 points), normal kidney function and GBS-ADL (0 points) and a Charlson index score of 5 (16 points). The total score would be 3+6+11+8+0+16 = 44 points, placing this patient well within risk group 1. If this patient had all other variables constant but a functional decline, with a GBS-ADL score of 7, this would result in a total score of 44+30 = 74, placing the patient in risk group 3. The risk attributed to the functional decline would be equivalent to a hemoglobin drop from 98 to 55 g/L. Would it infer the same sense of urgency to the clinician?

References - statistical appendix

1. *R: A language and environment for statistical computing.* [computer program]. Version. VIenna; 2014.
2. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* Feb 28 1996;15(4):361-387.
3. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol.* Oct 2006;59(10):1087-1091.
4. Horton NJ, Kleinman KP. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *Am Stat.* Feb 2007;61(1):79-90.
5. Janssen KJ, Donders AR, Harrell FE, Jr., et al. Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol.* Jul 2010;63(7):721-727.
6. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Med Res Methodol.* 2010;10:7.
7. Moons KG, Donders RA, Stijnen T, Harrell FE, Jr. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol.* Oct 2006;59(10):1092-1101.
8. Steyerberg EW. *Clinical Prediction Models - A Practical Approach to Development, Validation, and Updating*. New York: Springer; 2009.
9. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med.* Jan 15 2006;25(1):127-141.
10. Harrell FE, Jr., Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst.* Oct 5 1988;80(15):1198-1202.
11. Harrell Jr FE. *Regression Modeling Strategies - With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer-Verlag; 2001.
12. O'Brien RM. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity* 2007-10-01 2007;41(5):673-690.
13. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med.* May 10 2011;30(10):1105-1117.
14. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med.* Jan 30 2008;27(2):157-172; discussion 207-112.
15. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med.* Jan 15 2011;30(1):11-21.
16. Pencina MJ, D'Agostino RB, Pencina KM, Janssens AC, Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol.* Sep 15 2012;176(6):473-481.

17. Chambless LE, Cummiskey CP, Cui G. Several methods to assess improvement in risk prediction models: extension to survival analysis. *Stat Med.* Jan 15 2011;30(1):22-38.

18. Uno H, Tian L, Cai T, Kohane IS, Wei LJ. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Stat Med.* Jun 30 2013;32(14):2430-2442.

19. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med.* Aug 30 2004;23(16):2567-2586.

20. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* Aug 2001;54(8):774-781.

21. Tibshirani R. Regression shrinkage and selection via the Lasso. *J. Royal. Statist. Soc B.* 1996;58(1):267-288.

22. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med.* Feb 28 1997;16(4):385-395.

```
##########################################################
##                                                      ##
##            Appendix 2 - R code                       ##
##                                                      ##
##########################################################

## Data import from SPSS   ##

> library(rms)
> library(Hmisc)
> data <- spss.get() #file path in brackets

## Basic properties, table 1 (article) ##

> names(data)

"Nr"          "age"         "sex"         "status"
"time"        "cci"         "dressing"    "eating"
"physical"    "spontaneous" "hygiene"     "toilet"
"adl"         "hb"          "gfr"         "albumin"
"bnp"         "bmi"         "control"

> describe(data)

##########################################################
## 1. outcome                                           ##
##                                                      ##
##########################################################

## Determining follow-up and censoring ##

> library (survival)
> S <- Surv(time, status)
> cens.time <- ifelse(status == ”alive”, time, NA)
summary(cens.time)

## Baseline survival plot, figure 1 (appendix 1) ##

> S.years <- Surv(time/365.25, status)

> survplot(npsurv(S.years~1), xlab = "years", xlim =
c(0,4), time.inc = 1, lwd = 1.5, n.risk = T, y.n.risk =
0.05, cex.n.risk = 1, adj.n.risk = 0.5)

## discarding cases with missing outcome ##

> missing.outcome <- is.na(data$status)
> data <- data[missing.outcome == FALSE,]
> attach(data)

##########################################################
## 2. Crude analysis                                    ##
##                                                      ##
##########################################################
```

```
## Crude analysis for all predictors, table 2 (article)##

> summary(coxph(S~age))
> summary(coxph(S~sex))
> summary(coxph(S~cci))
> summary(coxph(S~dressing))
> summary(coxph(S~eating))
> summary(coxph(S~physical))
> summary(coxph(S~spontaneous))
> summary(coxph(S~hygiene))
> summary(coxph(S~toilet))
> summary(coxph(S~adl))
> summary(coxph(S~hb))
> summary(coxph(S~gfr))
> summary(coxph(S~albumin))
> summary(coxph(S~bnp))
> summary(coxph(S~bmi))
> summary(coxph(S~control))

## Discarding the separate ADL items ##

> data <- data[,-c(7:12)]

############################################################
## 3. Missing data                                        ##
##                                                        ##
############################################################

## Defining covariates ##

> covs <- data[, c("age", "sex", "cci", "adl", "hb",
"clearance", "albumin", "bnp", "bmi", ”control”)]

## Plotting missing, figure 2 (appendix 1) ##

> missing <- naclus(covs)
> naplot(missing, which = ”na per var”)

## Discarding the BNP variable ##

> data <- data[,-8]

## Determining associations with missing albumin ##

> missing.albumin <- ifelse(is.na(albumin), 1, 0)
> lrm(missing.albumin~age+sex+cci+clearance+bmi+adl+hb)
> chisq.test(missing.albumin, sex)
> oneway.test(cci~missing.albumin)
> tapply (cci, missing.albumin, mean)

## Creating a transcan object for imputation ##

> trans <-transcan(~age + sex + cci + adl + hb +
clearance + albumin + bmi + control, imputed = T, data =
data)
```

```
## Imputing albumin, gfr and bmi ##

> albumin.imputed <- as.integer(impute(trans, albumin,
data = data))
> gfr.imputed <- as.integer(impute(trans, gfr, data =
data))
> bmi.imputed <- as.integer(impute(trans, bmi, data =
data))

## Testing imputed variables, table 1 (appendix 1) ##

> summary(coxph(S~albumin.imputed)
> summary(coxph(S~gfr.imputed)
> summary(coxph(S~bmi.imputed)

## Imputed variables put into original dataset ##

> data$gfr <- gfr.imputed
> data$albumin <- albumin.imputed
> data$bmi <- bmi.1
> attach(data)

############################################################
## 4. Variable considerations                             ##
##                                                        ##
############################################################

## Screening for outliers, figure 3 (appendix 1) ##

> par(mfrow = c(3,3))
> boxplot(age)
> boxplot(cci)
> boxplot(hb)
> boxplot(gfr)
> boxplot(albumin)
> boxplot(bmi)
> boxplot(adl)

## Winsorising at 99th percentile, table 2(appendix 1) ##

> describe(gfr)
> gfr.winsorised <- ifelse(gfr > 118, 118, gfr)
> summary(coxph(S~gfr.winsorised))
> data$gfr <- gfr.winsorised
> attach(data)

############################################################
## Transformations for haemoglobin figure 4, appendix 1##
############################################################

## Range is obtained ##

> describe(hb)

## Linear model fitted and plotted ##
```

```
> hbfit.linear <- cph(S~hb, data = data)
> plot(Predict(hbfit.linear, hb = seq(53, 179, by = 1)),
xlab = "Haemoglobin g/L" , anova = anova(hbfit.linear),
pval = T, data = llist(hb))

## Dichotomous model##

> data$anemia <- ifelse(sex == ”male”, ifelse(hb < 130,
”yes”, ”no”), ifelse (hb < 120, ”yes”, ”no”))
> dd <- datadist (data)
> options(datadist = ”dd”)
> hbfit.dichotomous <- cph(S~anemia, data = data)
> plot(Predict(hbfit.dichotomous), xlab = "Anemia", anova
= anova(hbfit.dichotomous), pval = T)

## Categorical model ##

> data$hbcat <- ifelse(hb < 92.25, 1, ifelse(hb < 118, 2,
ifelse(hb < 130, 3, ifelse (hb < 148, 4, 5))))
> dd <- datadist (data)
> options(datadist = ”dd”)
> hbfit.categorical <- cph(S~as.factor(hbcat), data =
data)
> plot(Predict(hbfit.categorical), anova =
anova(hbfit.categorical), xlab = ”Haemoglobin g/L”, pval
= T)

## Restricted cubic spline ##

> hbfit.spline <- cph(S~rcs(hb,4), data = data)
> plot(Predict(hbfit.spline, hb = seq(53,179, by = 1)),
anova = anova(hbfit.spline), pval = T, xlab =
"Haemoglobin g/L", data = llist(hb))

## Testing other continuous variables ##

> agefit <- cph(S~rcs(age,4), data = data)
> anova(agefit)

> ccifit <- cph(S~rcs(cci,4), data = data)
> anova(ccifit)

> albuminfit <- cph(S~rcs(albumin,4), data = data)
> anova(albuminfit)

> bmifit <- cph(S~rcs(bmi,4), data = data)
> anova(bmifit)

> gfrfit <- cph(S~rcs(gfr,4), data = data)
> anova(gfrfit)

> adlfit <- cph(S~rcs(adl,4), data = data)
> anova(adlfit)
```

```
##############################################################
## Transformations for ADL figure 5, appendix 1             ##
##############################################################

> describe(adl)

## Linear fit ##

> adlfit.linear <- cph(S~adl, data = data)
> plot(Predict(adlfit.linear, adl = seq(0,25, by = 1)),
anova = anova(adlfit.linear), pval = T, data =
llist(adl), xlab = "GBS-ADL")

## Dichotomised at median ##

> data$adl.dichotomised <- ifelse(adl<5,0,1)
> dd <- datadist(data)
> options(datadist = "dd")
> adlfit.dichotomous <- cph(S~adl.dichotomised, data =
data)
> plot(Predict(adlfit.dichotomous), anova =
anova(adlfit.dichotomous), pval = T)

## Categorised at quartiles ##

> data$adl.quartiles <- ifelse(adl<3, 1, ifelse(adl<6,2,
ifelse(adl<10, 3,4)))
> dd <- datadist(data)
> options(datadist = "dd")
> adlfit.quartiles <- cph(S~as.factor(adl.quartiles),
data = data)
> plot(Predict(adlfit.quartiles), anova =
anova(adlfit.quartiles), pval = T)

## Two-degree polynomial ##

> adlfit.poly <- cph(S~pol(adl,2), data = data)
> plot(Predict(adlfit.poly, adl = seq(0,25,by=1)), anova
= anova(adlfit.poly), pval = T, xlab = "GBS-ADL", data =
llist(adl))

## four-knot spline ##

> adlfit.spline <- cph(S~rcs(adl,4), data = data)
> plot(Predict(adlfit.spline, adl = seq(0,25,by=1)),
anova = anova(adlfit.spline), pval = T, xlab = "GBS-ADL",
data = llist(adl))

## Log fit ##

> adlfit.log <- cph(S~log(adl+1), data = data)
> plot(Predict(adlfit.log, adl = seq(0,25,by=1)), anova =
anova(adlfit.log), pval = T, xlab = "GBS-ADL", data =
llist(adl))
```

```
#############################################################
## 5. Fitting the multivariate models                      ##
##                                                         ##
#############################################################

## The model without ADL ##

> model1 <- cph(S~age + sex + cci + rcs (hb, 4) + albumin
+ bmi + control + gfr, x = T, y = T, surv = T, data =
data)

## The full model ##

> model2 <- cph(S~age + sex + cci + rcs(hb, 4) + albumin
+ bmi + control + gfr + log (adl + 1), x = T, y = T, surv
= T, data = data)

#############################################################
## 6. Multicolinearity                                     ##
##                                                         ##
#############################################################

> cor(data)
> vif(model1)
> vif(model2)

#############################################################
## 7. Interactions                                         ##
##                                                         ##
#############################################################

## Global tests for the model without ADL ##

> z1 <- predict(model1, type = ”terms”)

> age.ia <- z1[,”age”]
> all.others <- z1[,-1]
> anova(cph(S~age.ia*all.others))

> sex.ia <- z1[,”sex”]
> all.others <- z1[,-2]
> anova(cph(S~sex.ia*all.others))
> cph(S~sex.ia*all.others)

> cci.ia <- z1[,"cci"]
> all.others <- z1[,-3]
> anova(cph(S~cci.ia*all.others))

> hb.ia <- z1[,"hb"]
> all.others <- z1[,-4]
> anova(cph(S~hb.ia*all.others))

> albumin.ia <- z1[,"albumin"]
> all.others <- z1[,-5]
> anova(cph(S~albumin.ia*all.others))
```

```
> bmi.ia <- z1[,"bmi"]
> all.others <- z1[,-6]
> anova(cph(S~bmi.ia*all.others))
> cph(S~bmi.ia*all.others)

> control.ia <- z1[,"control"]
> all.others <- z1[,-7]
> anova(cph(S~control.ia*all.others))

> gfr.ia <- z1[,"gfr"]
> all.others <- z1[,-8]
> anova(cph(S~gfr.ia*all.others))

## The full model ##

> z2 <- predict(model2, type = ”terms”)

> age.ia2 <- z2[,"age"]
> all.others.2 <- z2[,-1]
> anova(cph(S~age.ia2*all.others.2))

> sex.ia2 <- z2[,"sex"]
> all.others.2 <- z2[,-2]
> anova(cph(S~sex.ia2*all.others.2))

> cci.ia2 <- z2[,"cci"]
> all.others.2 <- z2[,-3]
> anova(cph(S~cci.ia2*all.others.2))

> hb.ia2 <- z2[,"hb"]
> all.others.2 <- z2[,-4]
> anova(cph(S~hb.ia2*all.others.2))

> albumin.ia2 <- z2[,"albumin"]
> all.others.2 <- z2[,-5]
> anova(cph(S~albumin.ia2*all.others.2))

> bmi.ia2 <- z2[,"bmi"]
> all.others.2 <- z2[,-6]
> anova(cph(S~bmi.ia2*all.others.2))

> control.ia2 <- z2[,"control"]
> all.others.2 <- z2[,-7]
> anova(cph(S~control.ia2*all.others.2))

> gfr.ia2 <- z2[,"gfr"]
> all.others.2 <- z2[,-8]
> anova(cph(S~gfr.ia2*all.others.2))
> cph(S~gfr.ia2*all.others.2)

> adl.ia <- z2[,"adl"]
> all.others <- z2[,-9]
> anova(cph(S~adl.ia*all.others))
> cph(S~adl.ia*all.others)
```

```
## Updating the models with the interactions ##

> model1 <- cph(S~age + sex * bmi + cci + rcs (hb, 4) +
albumin + control + gfr, data = data, x = T, y = T, surv
= T)

> model2 <- cph(S~age + sex * bmi + cci + rcs (hb, 4) +
albumin + control + gfr * log (adl + 1), data = data, x =
T, y = T, surv = T)

##########################################################
## 8. Proportional hazards assumption                    ##
##                                                       ##
##########################################################

> z3 <- predict(model1, type = "terms")
> model1.short <- cph(S~z3, x = T, y = T)
> ph1 <- cox.zph(model1.short, transform = "identity")
> ph1

> z4 <- predict(model2, type = ”terms”)
> model2.short <- cph(S~z4, x = T, y = T)
> ph2 <- cox.zph(model2.short, transform = ”identity”)
> ph2
> plot(ph2, var = ”gfr”) ##figure 6 (appendix 1) ##

##########################################################
## 9. Influential observations                           ##
##                                                       ##
##########################################################

> inf1 <- which.influence(model1)
> show.influence(inf1, dframe = data)

> inf2 <- which.influence(model2)
> show.influence(inf2, dframe = data)
> inf2

## Sensitivity analysis without influential for ADL ##

> subset <- data[-c(3,25,38,56,67,69,95,108,161),]
> attach(subset)
> S.sens <- Surv (time, status)
> sensitivity.model <- cph(S.sens~age + sex * bmi + cci+
rcs (hb, 4) + albumin + control + gfr * log(adl + 1), x=
T, y = T, surv = T, data = subset)
> sensitivity.model
> anova(sensitivity.model)
> detach(subset)
```

```
#############################################################
## 10. Relative contribution of ADL, figure 1(article) ##
##     figure 7(appendix)                                  ##
##                                                         ##
#############################################################

> plot(anova(model1), margin = "P", rm.ia = TRUE)
> plot(Predict(model1), anova = anova(model1), pval = T)

> plot(anova(model2), margin = "P", rm.ia = TRUE)
> plot(Predict(model2), anova = anova(model1), pval = T)

#############################################################
## 11. Added value of ADL, table 3 (article)              ##
##                                                         ##
#############################################################

## Likelihood ratio χ² test ##

> lrtest(model1, model2)

## Discrimination ##

> library (survC1)
> mydata <- as.matrix(data[,c("time", "status")])
> Inf.Cval.Delta(mydata, model1$x, model2$x, tau = 1428)

## NRI>0 and IDI ##

> library(survIDINRI)
> i <- IDI.INF(mydata, model1$x, model2$x, t0 = 1428)
> IDI.INF.OUT(i)

#############################################################
## 12. Internal validation                                ##
##                                                         ##
#############################################################

> validate(model1, B = 1000)
> validate(model2, B = 1000)

#############################################################
##                                                         ##
## 13. Updating the model                                 ##
##                                                         ##
#############################################################

>library(glmpath)
> mydata <- list(x = predict(model2, type = "ccterms"),
time = data$time, status = data$status)
> path <- coxpath(data = mydata)

##creating figure 8 appendix ##
```

```
> plot(path)
> plot(path, type = ”aic”)

## Determining the shrinkage factors ##

> lasso.factors <- path$b.predictor[path$aic ==
min(path$aic),]

## Shrinking the lasso.coefs ##

> lasso.coefs <- model2$coef

> lasso.coefs[”age”] <- lasso.coefs[”age”] *
lasso.factors[1]
> lasso.coefs[”sex”] <- lasso.coefs[”sex”] *
lasso.factors[2]
> lasso.coefs[”bmi”] <- lasso.coefs[”bmi”] *
lasso.factors[2]
> lasso.coefs[”cci”] <- lasso.coefs[”cci”] *
lasso.factors[3]
> lasso.coefs[”hb”] <- lasso.coefs[”hb”] *
lasso.factors[4]
> lasso.coefs[”hb’”] <- lasso.coefs[”hb’”] *
lasso.factors[4]
> lasso.coefs[”hb’’”] <- lasso.coefs[”hb’’”] *
lasso.factors[4]
> lasso.coefs[”albumin”] <- lasso.coefs[”albumin”] *
lasso.factors[5]
> lasso.coefs[”control”] <- lasso.coefs[”control”] * 0
> lasso.coefs[”gfr”] <- lasso.coefs[”gfr”] *
lasso.factors[7]
> lasso.coefs[”adl”] <- lasso.coefs[”adl”] *
lasso.factors[7]
> lasso.coefs[”sex * bmi”] <- lasso.coefs[”sex * bmi”] *
lasso.factors[2] * lasso.factors[2]
> lasso.coefs[”gfr * adl”] <- lasso.coefs[”gfr * adl”] *
lasso.factors[7] * lasso.factors[7]

## Updating the model ””

> lassomodel <- model2
> lassomodel$coefficients <- lasso.coefs

## Plotting nomogram, figure 2 (article) ##

> plot(nomogram(lassomodel, age = c(60,80,100), albumin =
c(15,20,30,40,45), bmi = c(15,20,25,30,35), hb =
c(50,70,90,110,150,175), interact = list(gfr =
c(27,36,51), adl, bmi, sex), lp = T, lp.at = c(-4,-
2,0,2), nint = 5, maxscale = 50))

## creating four risk groups ##
```

```
> risk.group <- cut2(as.numeric
(lassomodel$linear.predictor), g = 4)
> levels(risk.group) <- as.character(1:4)

## Kaplan-Meier plot, figure 3 (article)

> survplot(npsurv(S~risk.group, data = data), xlim =
c(0,1318), label.curves = FALSE, conf = "none", n.risk =
T, xlab = "follow-up (days)", cex.nrisk = 0.8, ylab =
"Fraction survivors", time.inc = 364, sep.n.risk = 0.03,
y.n.risk = 0, col = c(1,2,3,4), lty = 1)


### END ###
```

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract *"Cohort study" in title and abstract, page 1 and page 2*
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found *results section in abstract includes the most important findings, related to the objectives, page 2.*
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported *Rationale of improved mortality prediction [introduction, 1st paragraph] The need for improvement in studies regarding ADL and mortality [Introduction, 2nd paragraph*], page 4
Objectives	3	State specific objectives, including any prespecified hypotheses *"we aim to determine the relative importance and added value of this ADL measurement compared to clinical data, with regard to mortality prediction" [introduction 3rd paragraph], page 4*
Methods		
Study design	4	Present key elements of study design early in the paper *The methods section starts with a general description [Methods 1st paragraph], page 4. The1st paragraph of the statistical method section also describes the study design, page 6*
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection *setting, location, dates, follow-up and data collection are described on pages 4 to 6. Exposure was not dichotomous but continuous [Methods]*
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *eligibility, and methods for selection of participants [Methods, page 5] Follow-up [statistical method, page 7]*
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Not matched, Not applicable*
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable *all variables are presented in methods section [Methods, page 5 and 6] Variable transformations are presented in statistical methods section, page 7*
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group *sources of data and details of assessment are described. [Methods, page 5 and 6]*
Bias	9	Describe any efforts to address potential sources of bias *selection bias was addressed by including the group allocation variable in all analyses, as mentioned in discussion, page 13.*
Study size	10	Explain how the study size was arrived at

		"As a secondary analysis, no specific power calculation was done " [statistical methods, page 6] Before the original study a power calculation was performed, suggesting a needed sample size of 202 patients, albeit with a completely different study question.
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why *described in general in paper [statistical methods, page 6 to 9] and in detail in appendix 1*
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding *described in general in paper [statistical methods, page 6 to 9] and in detail in appendix 1*
		(*b*) Describe any methods used to examine subgroups and interactions *described in general in paper [statistical methods, page 7] and in detail in appendix 1*
		(*c*) Explain how missing data were addressed *described in general in paper [statistical methods, page 7] and in detail in appendix 1*
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *described in general in paper [results, page 9] and in detail in appendix 1*
		(*e*) Describe any sensitivity analyses *described in general in paper [statistical methods, page 7] and in detail in appendix 1*

Continued on next page

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed *eligibility is described in methods section and further details are found in the referenced original paper [methods page 5, reference 26]*
		(b) Give reasons for non-participation at each stage *This is described in the methods section and, in greater detail, in the referenced paper [methods page 5, reference 26]*
		(c) Consider use of a flow diagram *a reference to the flowchart in reference 26 is found in the methods section [methods page 5, reference 26]*
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders *descriptive data is found in table 1, page 9*
		(b) Indicate number of participants with missing data for each variable of interest *This is described in table 1, page 9 in the text of the results section, page 10 and in appendix 1.*
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount) *Follow-up time is summarised in results section, page 9 and appendix 1*
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time *number of events is reported in results section, page 9.*
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included *unadjusted estimates are presented in table 2 page 10. Multivariate estimates in figure 1, figure 2 and appendix 1.*
		(*b*) Report category boundaries when continuous variables were categorized *no categorisation or dichotomisation was done.*
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period *risks were displayed using a nomogram, figure 2*
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses *interactions are reported in statistical methods page 7, results page 11, appendix 1. One sensitivity analysis was done, with a summary in results page 11, larger detail in appendix 1.*
Discussion		
Key results	18	Summarise key results with reference to study objectives *[discussion 1st paragraph, page 12]*
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias *selection bias and Overfitting adressed in discussion page 12 and 13 and appendix 1 [methods].*
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence

		objectives, limitations, number of analysis and similar studies are discussed in discussion section, page 13.
Generalisability	21	Discuss the generalisability (external validity) of the study results *Validity (external and internal) is discussed in discussion section, page 13 and appendix 1.*
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based *Sources of funding, and roles of these, are presented in "funders", page 15.*

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract *"Cohort study" in title and abstract, page 1 and page 2*
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found *results section in abstract includes the most important findings, related to the objectives, page 2.*
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported *Rationale of improved mortality prediction [introduction, 1st paragraph] The need for improvement in studies regarding ADL and mortality [Introduction, 2nd paragraph]*, page 4
Objectives	3	State specific objectives, including any prespecified hypotheses *"we aim to determine the relative importance and added value of this ADL measurement compared to clinical data, with regard to mortality prediction" [introduction 3rd paragraph], page 4*
Methods		
Study design	4	Present key elements of study design early in the paper *The methods section starts with a general description [Methods 1st paragraph], page 4. The1st paragraph of the statistical method section also describes the study design, page 6*
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection *setting, location, dates, follow-up and data collection are described on pages 4 to 6. Exposure was not dichotomous but continuous [Methods]*
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *eligibility, and methods for selection of participants [Methods, page 5] Follow-up [statistical method, page 7]*
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Not matched, Not applicable*
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable *all variables are presented in methods section [Methods, page 5 and 6] Variable transformations are presented in statistical methods section, page 7*
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group *sources of data and details of assessment are described. [Methods, page 5 and 6]*
Bias	9	Describe any efforts to address potential sources of bias *selection bias was addressed by including the group allocation variable in all analyses, as mentioned in discussion, page 13.*
Study size	10	Explain how the study size was arrived at

		"As a secondary analysis, no specific power calculation was done " [statistical methods, page 6] Before the original study a power calculation was performed, suggesting a needed sample size of 202 patients, albeit with a completely different study question.
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why *described in general in paper [statistical methods, page 6 to 9] and in detail in appendix 1*
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding *described in general in paper [statistical methods, page 6 to 9] and in detail in appendix 1*
		(*b*) Describe any methods used to examine subgroups and interactions *described in general in paper [statistical methods, page 7] and in detail in appendix 1*
		(*c*) Explain how missing data were addressed *described in general in paper [statistical methods, page 7] and in detail in appendix 1*
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *described in general in paper [results, page 9] and in detail in appendix 1*
		(*e*) Describe any sensitivity analyses *described in general in paper [statistical methods, page 7] and in detail in appendix 1*

Continued on next page

Formatted: Highlight

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed *eligibility is described in methods section and further details are found in the referenced original paper [methods page 5, reference 26]*
		(b) Give reasons for non-participation at each stage *This is described in the methods section and, in greater detail, in the referenced paper [methods page 5, reference 26]*
		(c) Consider use of a flow diagram *a reference to the flowchart in reference 26 is found in the methods section [methods page 5, reference 26]*
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders *descriptive data is found in table 1, page 9*
		(b) Indicate number of participants with missing data for each variable of interest *This is described in table 1, page 9 in the text of the results section, page 10 and in appendix 1.*
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount) *Follow-up time is summarised in results section, page 9 and appendix 1*
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time *number of events is reported in results section, page 9.*
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included *unadjusted estimates are presented in table 2 page 10. Multivariate estimates in figure 1, figure 2 and appendix 1.*
		(*b*) Report category boundaries when continuous variables were categorized *no categorisation or dichotomisation was done.*
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period *risks were displayed using a nomogram, figure 2*
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses *interactions are reported in statistical methods page 7, results page 11, appendix 1. One sensitivity analysis was done, with a summary in results page 11, larger detail in appendix 1.*
Discussion		
Key results	18	Summarise key results with reference to study objectives *[discussion 1st paragraph, page 12]*
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias *selection bias and Overfitting adressed in discussion page 12 and 13 and appendix 1 [methods].*
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence

		objectives, limitations, number of analysis and similar studies are discussed in discussion section, page 13.
Generalisability	21	Discuss the generalisability (external validity) of the study results *Validity (external and internal) is discussed in discussion section, page 13 and appendix 1.*
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based *Sources of funding, and roles of these, are presented in "funders", page 15.*

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Importance and added value of functional impairment in order to predict mortality: a cohort study in swedish medical inpatients

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-014464.R2
Article Type:	Research
Date Submitted by the Author:	27-Feb-2017
Complete List of Authors:	Torisson, Gustav; Lunds Universitet, Clinical Memory Research Unit Dept of Clinical Sciences Stavenow, Lars; Skanes universitetssjukhus Malmo, Department of Internal Medicine Minthon, Lennart; Clinical Memory Research Unit, Department of Clinical Sciences Londos, Elisabet; Clinical Memory Research Unit, Department of Clinical Sciences, Lund University
Primary Subject Heading:	Geriatric medicine
Secondary Subject Heading:	Research methods
Keywords:	GERIATRIC MEDICINE, STATISTICS & RESEARCH METHODS, Delirium & cognitive disorders < PSYCHIATRY

Importance and added value of functional impairment in order to predict mortality: a cohort study in swedish medical inpatients

Authors

Gustav Torisson[1], PhD, Gustav.Torisson@med.lu.se

Lars Stavenow[2], associate professor, Lars.E.Stavenow@skane.se

Lennart Minthon[1], professor, Lennart.Minthon@skane.se

Elisabet Londos[1], associate professor, Elisabet.Londos@skane.se

Affiliations

[1] Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden.

[2] Department of Internal Medicine, Skåne University Hospital, Malmö Sweden.

Corresponding author:

Gustav Torisson, Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden.

Adress: Simrisbanvägen 14, 20502 Malmö.

E-mail: gustav.torisson@med.lu.se.

Telephone: +46 40 33 77 85

Word count: 2857

Abstract

Background: Accurate estimation of prognosis in multimorbid hospital patients could improve quality of care. This study aims to determine the relative importance and added value of a performance-based ADL (activities of daily living) measure with regard to mortality prediction.

Methods: Two hundred inpatients, aged over 60 years, were recruited at the Department of General Internal Medicine at a tertiary university hospital. Two nested survival models were built, one with established risk factors (age, sex, Charlson comorbidity index, haemoglobin, albumin, body mass index, and glomerular filtration rate)**,** and one using the same covariates with the GBS (Gottfries-Bråne-Steen) ADL measure added. The relative importance of GBS-ADL was evaluated in the full model. The added value of GBS-ADL was determined by comparing the nested models using four approaches: difference in overall χ^2, discrimination, continuous net reclassification index (NRI > 0) and integrated discrimination improvement (IDI).

Results: In the full model, GBS-ADL was the single most important predictor of mortality (χ^2 - df = 30, p <0.001). The likelihood ratio χ^2 test showed significant added value of ADL (p<0.001). The c statistic was 0.78 with ADL and 0.72 without, (difference 0.058, 95% CI = 0.022 to 0.094). The NRI > 0 was 0.42 (95% CI 0.20 to 0.58) and IDI 0.15 (95% CI 0.07 to 0.22).

Conclusions: Compared to a set of available clinical risk factors, impairment in ADL was a stronger predictor of all-cause mortality, showing substantial added value. Implementing quantitative ADL measurements could enable more appropriate and individual care for the elderly.

Keywords: aging, comorbidity, mortality, functional status, statistical modeling

Strengths and limitations of this study

- A rigorous survival analysis was used to determine the relative importance of impaired ADL, compared to readily available clinical information.
- Four different methods were used to determine the added prognostic value of impaired ADL.
- However, the study was a secondary analysis, using data from an intervention study, and a larger study is needed.
- Only one ADL measurement was used, the results need to be confirmed for other ADL scales to be considered generalisable.

INTRODUCTION

Improving the accuracy of prognostic estimates could have several benefits for medical inpatients. Such benefits include reduced overtreatment, such as polypharmacy or the use of life-sustaining measures inconsistently with patients' preferences[1-4]. Other elderly patients are withheld treatment due to an incorrectly supposed poor prognosis, this could possibly be another important aspect[5-7]. Furthermore, patients with poor prognosis may prefer improved quality of life over extended survival. Therefore, accurate estimates could support doctors initiating a discussion regarding goals of care[8]. In addition, advance care planning could help patients and families to make necessary arrangements and increase quality of life[9-11].

Impairment in ADL (activities of daily living) is a well-known predictor of mortality and lower quality of life in hospitalized and community-dwelling elderly[12-20]. However, the majority of studies use interview-based scales[13, 15, 18], shown to differ significantly from performance-based ones[21, 22]. In addition, several studies use regression models without reporting overall performance[14, 15, 18, 23] and only few studies determine the added value of ADL[13, 15]. Recently, novel statistical methods have been introduced to establish the incremental value of prognostic markers[24].

In the present study, we aim to use these methods in order to determine the relative importance and added value of a performance-based ADL measure compared to clinical data, with regard to mortality prediction.

METHOD

This study constitutes a secondary analysis, all patients were concurrently taking part in a prospective trial, aiming to improve quality of care[25].

Setting

The study was carried out at the Department of General Internal Medicine at Skåne University Hospital in Malmö, Sweden. This teaching hospital provides care to the city's approximately 300,000 inhabitants. The department has four wards, with a total of 100 beds. Patients are admitted through the hospital's Emergency Department. Normally, the patients in the

department are elderly with multiple comorbidities. More specialized medical departments (cardiology, nephrology, endocrinology etc.) are separate and were not included in this study.

Patients

The recruitment of patients, that took place in 2009 and 2010, has been described in detail in a previous publication, including a flowchart[25]. In short, patients aged over 60 years, living in their own homes were eligible. Exclusion criteria comprised of terminal disease, language barrier, blindness/deafness/aphasia or other disease with inability to communicate, transfer to another department/ICU, early discharge and isolation due to communicable disease.

In total, 200 patients were included and underwent a baseline measurement. One half (101) of the patients constituted a control group while the other half (99) received a hospital-based, multidisciplinary intervention aiming to reduce rehospitalizations. The intervention consisted of a medication overview, improved discharge planning, telephone follow-up and improved liaison with GPs. Group allocation (intervention or control) used convenience sampling with geographic selection. At one-year follow-up, the intervention group had significantly fewer rehospitalizations than the control group[25].

ADL measurement

As part of the baseline measurement in the original trial, an ADL measurement was implemented by two experienced occupational therapists, who had received special training. The assessment was carried out when patients were stabilized, typically a few days into the admission.

The ADL subset of the GBS (Gottfries Bråne Steen) scale rates six items: dressing, food intake, physical activity, spontaneous activity, personal hygiene and toileting[26]. Items are scored on a performance-based 7-point scale ranging from 0 (best) to 6 (worst). For example, dressing is scored as follows:

0: Dresses and undresses without help

1:

2: Gets help with buttons, zips etc.

3:

4: Requires help from a caregiver to dress and undress but takes an active part

5:

6: Is completely dependent on a caregiver to be dressed and undressed

The points 1, 3 and 5 are not defined but are used by the observer to increase discrimination. Combining the six items gives a total ADL score of 0 (no impairment) to 36 (maximum impairment).

Other data from the original trial protocol

The Charlson comorbidity index was collected from the original protocol, to obtain a measure of combined comorbidity[27]. This index' performance concerning short-term and long-term mortality has recently been validated[28].

Data collection from medical records

Additional data was collected retrospectively regarding physiological and laboratory values. Since no blood samples were drawn in the original trial, only clinical data could be used. Candidate predictors were selected á priori on the basis of availability and previously established association with all-cause mortality. All data was obtained from the same hospital episode as ADL was measured. If a blood sample had not been drawn during that hospitalisation, the data point was labelled "missing". If several blood samples were taken during the hospitalisation, the one closest to admission was used. The following variables, all independently related to all-cause mortality, were collected: Body mass index (BMI), kg/m^2, Hemoglobin (Hb), g/L, estimated Glomerular filtration rate (eGFR), ml/min, Albumin, g/L, Brain Natriuretic Peptide (BNP), ηg/L[29, 30, 31-33].

Statistical method

The present study was a secondary analysis, thus no specific power calculation was done beforehand, this had been done for the original intervention study, albeit with a different research question {Torisson, 2013 #215}. The goal of the present study was to compare the GBS-ADL measurement with the best set of available clinical risk factors using survival analysis. First, we built a multivariate Cox regression model, called "model without ADL", using the established risk factors as covariates. Then, this model was refitted, with ADL added, to obtain the "full model". To determine the added value of ADL, the performance of these two models were compared. In addition, the relative importance of ADL was examined in the "full model".

The modelling algorithm is based on previous recommendations, primarily by Harrell et al and Steyerberg et al [24, 34-37]. All steps are explained in larger detail in Supplementary File 1. All modelling was executed in R, the script is supplied in Supplementary File 2.

1. **Outcome**. The study endpoint was mortality status on Feb 6th 2014. Follow-up was defined as time from discharge of the original hospitalisation.

2. **Crude analysis**. Separate bivariate proportional hazards regressions were carried out for all variables on their original scaling. Crude analysis were accomplished for all separate ADL items but in further analysis only the total GBS-ADL score was used.

3. **Missing data.** Missing values in covariates were quantified and controlled for systematic patterns resulting in their missing status. Missing values were then imputed using an imputational regression model.

4. **Variable transformations**: Haemoglobin was pre-specified to have a non-linear association with mortality. All other continuous variables were tested for non-linearity and transformed accordingly. Outliers were controlled for data entry errors and considered for truncation.

5. **Fitting the two multivariate models.** The "model without ADL" was fitted first, using the transformations and imputations described above. Then, ADL was added and the model was refitted to obtain the "full model".

6. **Multicolinearity.** The models were tested using the VIF (variance inflation factor).

7. **Interactions**: Pooled two-way interaction tests were carried out for all variables, in both models, separately. If the pooled test was significant, specific interactions were pursued for that variable.

8. **Proportional hazards**. The proportional hazards assumption was tested with global tests and Schoenfeld residual plots for each variable.

9. **Influential observations.** Observations with a standardised DfBeta > 0.20 standard errors were noted for each variable. As ADL was of particular interest, a sensitivity analysis was performed without this variable's influential observations.

10. **Determining the relative importance of ADL.** As the models contained non-linear variables as well as interactions, simple measurements of main effects, such as hazard

ratios, could not be used. To obtain an estimate of the relative importance of the different predictors, an ANOVA test was used instead, where interaction terms and non-linear terms are incorporated into each variable.

11. **Determining added value of ADL**. To determine added value, the "model without ADL" and the "full model" were compared using:

    a. **Likelihood ratio test.** Performed as a χ^2 testing the difference in Likelihood ratio between the models' χ^2 over df = number of additional independent variables.

    b. **Discrimination**, measured with the C, or concordance, statistic. The C statistic is the probability that, in a case-control pair, the case will be given a higher predicted risk from the model than the control. C statistics ranges from 0.5 (coin toss, useless) to 1.0 (perfect discrimination). The difference in C statistic between models was tested using the method described by Uno et al.[38] .

    c. **NRI >0 (Continuous net reclassification index)**[39, 40]. This index determines to what extent adding a new variable leads to a change in the correct direction of predicted risk for each observation (towards higher risk for deceased, towards lower for survivors). NRI ranges from 0 (no increased value, useless) to 1(all cases reclassified in the right direction). NRI has been shown to be more sensitive than change in C index, especially when the baseline model has a good performance.

    d. **IDI**. Integrated discrimination improvement. Originally developed by Pencina et al. for logistic models, IDI has been extended to time-to-event data[39, 41]. While NRI>0 measures the percentage of observations that have been reclassified, it cannot distinguish between a small change in prediction and a large. IDI, however, measures the mean amount of such change. IDI and NRI with confidence intervals were calculated with the method by Uno et al.[42]

12. **Internal validation**. Both models were internally validated through 1000 bootstrap resamples to estimate the amount of overfitting and to obtain optimism-corrected performance estimates.

13. **Updating and presenting final model**. The "full model" was updated through the use of a LASSO (least absolute shrinkage and selection operator) procedure to reduce the effects of overfitting[43, 44]. The updated LASSO model was used to build a nomogram, with which patients were stratified into four equally sized risk groups, displayed in a Kaplan-Meier graph.

RESULTS

In two cases, mortality status could not be obtained; these were discarded from further analysis. Of the remaining 198 cases, 126 were deceased at follow-up. The median follow-up time for survivors was 1428 days (range 1312-1548). Baseline characteristics are displayed in table 1.

Table 1. Baseline characteristics

Continuous variables	**mean (SD)**	**median (IQR)**	**min-max**
Age	83.4 (8.1)	85 (78-89)	60-100
Charlson comorbidity index	2.3 (1.5)	2 (1-3)	0-7
GBS-ADL, total	6.8 (5.7)	5 (2-10)	0-25
GBS-ADL, dressing	1.3 (1.4)	1(0-2)	0-5
GBS-ADL, food intake	0.1 (0.4)	0(0-0)	0-2
GBS-ADL, physical activity	2.0 (1.1)	2(2-2)	0-5
GBS-ADL, spontaneous activity	1.0 (1.2)	1(0-2)	0-5
GBS-ADL, hygiene	1.4 (1.4)	2(0-2)	0-5
GBS.ADL, toilet	0.9 (1.4)	0(0-1)	0-6
Hemoglobin, g/L	123 (19)	124 (111-136)	53-179
eGFR, ml/min, n = 197	42.3 (25)	37(26-51)	6-198
BMI, kg/m2, n = 195	24.7 (5.1)	24 (21-27)	14-42
Albumin, g/L, n = 181	31.5 (4.9)	32 (29-35)	14-42
BNP, ηg/L, n = 85	261 (297)	147 (54-377)	3-1618
Categorical variables	**number**	**percentage**	
Male sex	70	35%	
In intervention group in original study	99	50%	

Table 1. Baseline characteristics for the entire sample. n = 200 unless otherwise stated. ADL = activities of daily living, eGFR =estimated glomerular filtration rate, BMI = Body mass index, BNP = Brain natriuretic peptide.

The results from the crude analysis are presented in table 2.

Table 2. Crude analysis

Predictor	β	S.E	Wald X^2	p value	HR (95% CI)
GBS-ADL-total, points	0.08	0.013	37.8	<0.001	1.08 (1.06 - 1.11)
GBS-ADL-hygiene, points	0.38	0.06	37.7	<0.001	1.46 (1.29 - 1.65)
GBS-ADL-physical, points	0.46	0.08	36.0	<0.001	1.59 (1.36 - 1.84)
GBS-ADL-dressing, points	0.31	0.06	30.0	<0.001	1.36 (1.22 - 1.52)
eGFR, ml/min, n = 197	-0.029	0.005	29.3	<0.001	0.97 (0.96 - 0.98)
GBS-ADL-spontaneous, points	0.33	0.06	27.0	<0.001	1.40 (1.23 - 1.58)
Charlson index, points	0.22	0.06	15.2	<0.001	1.25 (1.18 - 1.40)
Hemoglobin, g/L	-0.019	0.005	14.6	<0.001	0.98 (0.97 - 0.99)
Albumin, g/L, n = 181	-0.064	0.018	13.1	<0.001	0.94 (0.90 - 0.97)
GBS-ADL- toileting, points	0.19	0.05	11.6	<0.001	1.20 (1.08 - 1.34)
Age, years	0.036	0.011	10.1	0.001	1.04 (1.01 - 1.06)
BMI, kg/m^2 , n = 195	-0.053	0.020	7.4	0.007	0.95 (0.91 - 0.99)
BNP, ηg/L, n = 85	0.0009	0.0003	6.7	0.01	1.001 (1 - 1.002)
ADL - food intake, points	0.34	0.21	2.7	0.10	1.41 (0.93 - 2.12)
Sex (0 = Female, 1 = Male)	0.29	0.18	2.6	0.11	1.34 (0.94 - 1.92)
Group in original study (0=control, 1=intervention)	0.11	0.18	0.37	0.54	1.12 (0.78 - 1.59)

Table 2. Crude Cox regression for all predictors, sorted by decreasing strength of association. S.E = standard error, HR = Hazard ratio, ADL = Activities of daily living, eGFR = glomerular filtration rate, BMI = body mass index, BNP = brain natriuretic peptide.

BNP was missing in 115 cases (58%) and the variable was discarded from further analysis. eGFR and BMI were missing in 1 and 3 cases, respectively; these were considered to be missing completely at random. Albumin was missing in 17 cases, these were predominantly female (15/17) and had lower scores on Charlson comorbidity index. Missing values were imputed with a minimal change in variable properties, see Supplementary File 1.

Hemoglobin was fitted using a 4-knot restricted spline and GBS-ADL was transformed using the natural logarithm. No other predictors showed significant non-linear properties and they were kept in their original form. eGFR had one extreme outlier at 198ml/min that was winsorized at the 99th percentile (118 ml//min).

A significant sex * BMI interaction was found and included into the models (low BMI was a significant predictor in men but not in women). Another interaction, eGFR * ADL, was included as well (ADL was a stronger predictor when eGFR was unimpaired and vice versa). No other significant interactions were found. No significant multicolinearities were found. The proportional hazards assumption was not violated. In the full model, 21 observations were influential, of which 9 for ADL and/or its interaction with eGFR. A sensitivity analysis with these cases removed showed a slight improvement in model fit and is presented in Supplementary File 1. However, all observations were kept in the models.

In the "full model", ADL was by far the most significant predictor. The relative importance of the predictor variables are shown in figure 1. All four measurements showed added value for model with ADL, see table 3.

Table 3. Added value of ADL

Model comparison	model without ADL	model with ADL	p value
Nagelkerke R^2	0.33	0.46	
Likelihood ratio χ^2	78.4 (11df)	121.0(13df)	<0.001
c statistic (95% CI)	0.72(0.67-0.76)	0.78(0.73-0.82)	0.001
½ NRI > 0 (95% CI)		0.42(0.20-0.58)	<0.001
IDI (95% CI)		0.15(0.07-0.22)	<0.001

Table 3. Comparison of the two nested survival models. NRI > 0 = continuous Net Reclassification Index, IDI = Integrated Discriminatory Improvement.

When bootstrapped 1000 times, the calibration slope of the "model without ADL" was 0.84 and of the "full model" 0.83. Optimism-corrected R^2 was 0.27 vs. 0.40, respectively. Optimism-corrected c statistics were 0.70 and 0.76. When the LASSO was employed to shrink coefficients and update the model, the mean shrinkage was 0.84. The nomogram using the updated model coefficients is shown in Supplementary File 1 and the subsequent Kaplan-Meier graph for the four risk groups are presented in figure 2.

DISCUSSION

In this study, we confirm that impaired ADL is an important predictor of mortality in elderly medical inpatients. The relative contribution of ADL was larger than of the available predictors in a real-life setting, including a comorbidity index, available physiological parameters and laboratory values. In addition, ADL showed a substantial added value when compared to a model combining all of these traditional predictors.

In the crude analysis, four of the GBS-ADL items were stronger predictors than the Charlson comorbidity index. Thus, a simple rating of dressing ability had better predictive value than a combined comorbidity measure, designed to predict mortality. This indicates that performance-based ADL measures are truly important mortality predictors in multimorbid patients. In multivariate analysis, impairment in ADL was by far the most important predictor and all four measures signaled added value when GBS-ADL was added to the traditional predictors.

The mechanism underlying the association between ADL and mortality is probably multifactorial. Impairment in ADL could contribute directly to mortality in some aspects. Obvious complications to functional decline include pressure sores, atrophy, falls, thrombosis etc. However, less intuitive factors could also apply, such as attaining multi-resistant bacteria or Clostridium Difficile[45 46]. Even more likely, ADL acts a proxy for a confounder not measured by the model. A possible such confounder is frailty, defined as an increased vulnerability, where small stressors lead to adverse outcomes, such as hospitalization or death[47]. The frailty phenotype includes unintentional weight loss, along with loss of strength, low physical activity, slow walking speed and exhaustion[48]. There is a considerable overlap between frailty, comorbidity and ADL impairment. Our study utilized specific measures for comorbidity and ADL impairment, but not for frailty. However, our model is most likely describing the effects of frailty as well.

Several methodological issues need to be addressed. First, the choice of ADL scale, where the GBS scale was chosen in order to facilitate implementation locally. There are large variations and lack of standardization regarding functional measures used in medical inpatients[49]. The GBS scale proved feasible and has been shown to have a good construct validity and inter-rater reliability[50]. In addition, the GBS-ADL has correlated strongly with other ADL measurements, for example Katz' index[51, 52]. Ideally, two different scales should have

been employed, to enable a comparison between scales and possibly improve generalisability. A potentially confounding issue was the concurrent non-randomized trial, i.e. the intervention could have affected mortality rates. However, the variable "control/intervention status" was included in all statistical analyses, both bivariate and multivariate, without any sign of bias. In addition, no power calculation was done, the sample size was small, and internal validation showed that our models were indeed overfitted, with a calibration slope of 0.83. This overfitting is probably not a result of having too many covariates but rather a result of the global interaction tests and tests of non-linearity. This multiple comparison situation has been called "testimation bias"[37]. The overall aim was not to develop the most comprehensive and parsimonious prediction model to use in future populations but to describe the importance and added value of ADL. Therefore, we prioritized not to miss clinically important interactions and/or transformations in the trade-off with overfitting. To compensate partly, we used a LASSO procedure to shrink estimates. The small sample size and the aim to compare ADL with the best possible model was also the reason underlying the imputation of missing values. In addition, the main diagnosis of the current hospitalisation was not included as a predictor in the analysis. The reason for this was the large heterogeneity of main diagnoses (with 97 different ICD codes in 200 patients), albeit this could possibly have been achieved with a larger sample size as well.

The primary strength of this study is the rigorous statistical approach. State-of-the-art methods were used in the model building to handle missing data, to address non-linearity, to screen for interactions, for model diagnostics and for internal validation. In addition, four different methods were applied to estimate added value. Previously, a study has showed increase in model χ^2 when adding a composite ADL measure, regarding 2-year mortality [15]. However, this study compared ADL only with comorbidity indices. With such a limited reference model it is likely that a new measure will add value but the final model could still perform poorly, which was reflected by low model χ^2 values and a final c statistic of 0.66. The use of comorbidity indices only as reference model is also far from the clinical reality. Another study shows increase in discrimination when adding an ADL measurement to a 1-year logistic regression mortality prediction model[13]. This study also starts with comorbidity indices alone and does not report any other measurement of overall performance (such as overall χ^2 or R^2). Our study compares ADL to a much more complex reference model and yet shows added value using both these previously applied measurements as well as several others.

Implications for further research include research regarding performance-based ADL scales, including the relation to specific frailty ratings. Larger studies could obtain head-to-head comparisons of ADL vs. disease-specific predictors, such as ejection fraction in heart failure.

Today, ADL is very often assessed in a variety of ways in medical inpatients, to assess the individuals' needs after discharge. Implementing a performance-based quantitative measurement could have many benefits, also apart from prognostic value, such as increased standardization and the possibility to follow a patient over time. As a final remark, mortality prediction is not all about avoiding overtreatment due to a poor prognosis. Our model identified 50 elderly multimorbid medical inpatients with a 90% chance of 3-year survival. This group should not be undertreated simply due to age discrimination.

In conclusion, an ADL measurement showed significant added value as a predictor of mortality in a multimorbid elderly hospital population. Implementation of standardized ADL measurements could lead to better prognostic estimates and in the end a more appropriate and individualised care for the elderly.

DECLARATIONS

List of abbreviations

ADL: activities of daily living

GBS: Gottfries-Bråne-Steen

BMI: body mass index

eGFR: estimated glomerular filtration rate

BNP: brain natriuretic peptide

NRI: net reclassification index

IDI: integrated discrimination improvement

Ethics approval

All patients enrolled in the original study gave written informed consent. Both the original trial and the secondary analysis have been approved by the regional ethics committee at Lund University.

Availability of data and materials

Since the participatns were not specifically asked for consent to share data, such sharing is not compatible with the current Swedish legislation. The data protection officer at Skåne University hospital, the data protection officer at Lund university as well as lawyers at the Swedish data protection authority have unanimously adviced us not to publish data, even if anonymized.

Conflicts of interests

The authors declare no financial relationships with any organizations that might have an interest in the submitted work and no other relationships or activities that could appear to have influenced the submitted work.

Funders

This study was financed by the Swedish Research Council (Vetenskapsrådet #523-2010-520), the Swedish Brain Power programme, the National Swedish Board of Health and Welfare and the Governmental Funding of Clinical Research within the National Health Services (ALF). All researchers acted independently to the funding bodies. The funding agencies had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Authors' contributions

The study was conceived by LS and LM. Data collection was done by GT, supervised by EL. GT performed all analysis and drafted the manuscript, which was critically revised by LS, LM and EL. All authors have approved the final version.

Acknowledgments

The authors would like to thand Sofia Raccuia, Jenny Cappelin and Anna Johansson for their assistance in the data acquisition.

REFERENCES

1. Frankl D, Oye RK, Bellamy PE. Attitudes of hospitalized patients toward life support: a survey of 200 medical inpatients. The American journal of medicine 1989;**86**(6):645-8
2. Heyland DK, Barwich D, Pichora D, et al. Failure to engage hospitalized elderly patients and their families in advance care planning. JAMA internal medicine 2013;**173**(9):778-87.
3. Somogyi-Zalud E, Zhong Z, Hamel MB, et al. The use of life-sustaining treatments in hospitalized persons aged 80 and older. Journal of the American Geriatrics Society 2002;**50**(5):930-4
4. Teno JM, Gozalo PL, Bynum JP, et al. Change in end-of-life care for Medicare beneficiaries: site of death, place of care, and health care transitions in 2000, 2005, and 2009. JAMA : the journal of the American Medical Association 2013;**309**(5):470-7.
5. Biganzoli L, Wildiers H, Oakman C, et al. Management of elderly patients with breast cancer: updated recommendations of the International Society of Geriatric Oncology (SIOG) and European Society of Breast Cancer Specialists (EUSOMA). The Lancet Oncology 2012;**13**(4):e148-60.
6. Tahir M, Robinson T, Stotter A. How not to neglect the care of elderly breast cancer patients? Breast 2011;**20**(4):293-6.
7. Wildiers H, Kunkler I, Biganzoli L, et al. Management of breast cancer in elderly individuals: recommendations of the International Society of Geriatric Oncology. The Lancet Oncology 2007;**8**(12):1101-15.
8. Dev S, Clare RM, Felker GM, et al. Link between decisions regarding resuscitation and preferences for quality over length of life with heart failure. European journal of heart failure 2012;**14**(1):45-53.
9. Steinhauser KE, Christakis NA, Clipp EC, et al. Factors considered important at the end of life by patients, family, physicians, and other care providers. JAMA : the journal of the American Medical Association 2000;**284**(19):2476-82
10. Detering KM, Hancock AD, Reade MC, et al. The impact of advance care planning on end of life care in elderly patients: randomised controlled trial. BMJ 2010;**340**:c1345.
11. Wright AA, Zhang B, Ray A, et al. Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. JAMA : the journal of the American Medical Association 2008;**300**(14):1665-73.
12. Han PK, Lee M, Reeve BB, et al. Development of a prognostic model for six-month mortality in older adults with declining health. Journal of pain and symptom management 2012;**43**(3):527-39.
13. Covinsky KE, Justice AC, Rosenthal GE, et al. Measuring prognosis and case mix in hospitalized elders. The importance of functional status. Journal of general internal medicine 1997;**12**(4):203-8
14. Fried LP, Kronmal RA, Newman AB, et al. Risk factors for 5-year mortality in older adults: the Cardiovascular Health Study. JAMA : the journal of the American Medical Association 1998;**279**(8):585-92
15. Inouye SK, Peduzzi PN, Robison JT, et al. Importance of functional measures in predicting mortality among older hospitalized patients. JAMA : the journal of the American Medical Association 1998;**279**(15):1187-93
16. Lee SJ, Lindquist K, Segal MR, et al. Development and validation of a prognostic index for 4-year mortality in older adults. JAMA : the journal of the American Medical Association 2006;**295**(7):801-8.

17. Teno JM, Harrell FE, Jr., Knaus W, et al. Prediction of survival for older hospitalized patients: the HELP survival model. Hospitalized Elderly Longitudinal Project. Journal of the American Geriatrics Society 2000;**48**(5 Suppl):S16-24
18. Walter LC, Brand RJ, Counsell SR, et al. Development and validation of a prognostic index for 1-year mortality in older adults after hospitalization. JAMA : the journal of the American Medical Association 2001;**285**(23):2987-94
19. Torisson G, Minthon L, Stavenow L, et al. Cognitive impairment is undetected in medical inpatients: a study of mortality and recognition amongst healthcare professionals. BMC geriatrics 2012;**12**:47.
20. Torisson G, Stavenow L, Minthon L, et al. Reliability, validity and clinical correlates of the Quality of Life in Alzheimer's disease (QoL-AD) scale in medical inpatients. Health and quality of life outcomes 2016;**14**:90.
21. Magaziner J, Zimmerman SI, Gruber-Baldini AL, et al. Proxy reporting in five areas of functional status. Comparison with self-reports and observations of performance. American journal of epidemiology 1997;**146**(5):418-28
22. Sherman SE, Reuben D. Measures of functional status in community-dwelling elders. Journal of general internal medicine 1998;**13**(12):817-23
23. Reuben DB, Rubenstein LV, Hirsch SH, et al. Value of functional status as a predictor of mortality: results of a prospective study. The American journal of medicine 1992;**93**(6):663-9
24. Steyerberg EW, Pencina MJ, Lingsma HF, et al. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. European journal of clinical investigation 2012;**42**(2):216-28.
25. Torisson G, Minthon L, Stavenow L, et al. Multidisciplinary intervention reducing readmissions in medical inpatients: a prospective, non-randomized study. Clinical interventions in aging 2013;**8**:1295-304.
26. Gottfries CG, Brane G, Gullberg B, et al. A new rating scale for dementia syndromes. Archives of gerontology and geriatrics 1982;**1**(4):311-30
27. Charlson M, Szatrowski TP, Peterson J, et al. Validation of a combined comorbidity index. Journal of clinical epidemiology 1994;**47**(11):1245-51
28. Frenkel WJ, Jongerius EJ, Mandjes-van Uitert MJ, et al. Validation of the Charlson Comorbidity Index in acutely hospitalized elderly adults: a prospective cohort study. Journal of the American Geriatrics Society 2014;**62**(2):342-6.
29. Janssen I, Katzmarzyk PT, Ross R. Body mass index is inversely related to mortality in older people after adjustment for waist circumference. Journal of the American Geriatrics Society 2005;**53**(12):2112-8.
30. Culleton BF, Manns BJ, Zhang J, et al. Impact of anemia on hospitalization and mortality in older adults. Blood 2006;**107**(10):3841-6.
31. Tonelli M, Wiebe N, Culleton B, et al. Chronic kidney disease and mortality risk: a systematic review. Journal of the American Society of Nephrology : JASN 2006;**17**(7):2034-47.
32. Goldwasser P, Feldman J. Association of serum albumin and mortality risk. Journal of clinical epidemiology 1997;**50**(6):693-703
33. Wallen T, Landahl S, Hedner T, et al. Brain natriuretic peptide predicts mortality in the elderly. Heart 1997;**77**(3):264-7
34. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine 1996;**15**(4):361-87.

35. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. European heart journal 2014;**35**(29):1925-31.
36. Harrell Jr FE. *Regression Modeling Strategies - With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer-Verlag, 2001.
37. Steyerberg EW. *Clinical Prediction Models - A Practical Approach to Development, Validation, and Updating*. New York: Springer, 2009.
38. Uno H, Cai T, Pencina MJ, et al. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Statistics in medicine 2011;**30**(10):1105-17.
39. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Statistics in medicine 2008;**27**(2):157-72; discussion 207-12.
40. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Statistics in medicine 2011;**30**(1):11-21.
41. Chambless LE, Cummiskey CP, Cui G. Several methods to assess improvement in risk prediction models: extension to survival analysis. Statistics in medicine 2011;**30**(1):22-38.
42. Uno H, Tian L, Cai T, et al. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. Statistics in medicine 2013;**32**(14):2430-42.
43. Tibshirani R. Regression shrinkage and selection via the Lasso. J Royal Statist Soc B 1996;**58**(1):267-88
44. Tibshirani R. The lasso method for variable selection in the Cox model. Statistics in medicine 1997;**16**(4):385-95
45. Chen TY, Anderson DJ, Chopra T, et al. Poor functional status is an independent predictor of surgical site infections due to methicillin-resistant Staphylococcus aureus in older adults. Journal of the American Geriatrics Society 2010;**58**(3):527-32.
46. Rao K, Micic D, Chenoweth E, et al. Poor functional status as a risk factor for severe Clostridium difficile infection in hospitalized older adults. Journal of the American Geriatrics Society 2013;**61**(10):1738-42.
47. Clegg A, Young J, Iliffe S, et al. Frailty in elderly people. Lancet 2013;**381**(9868):752-62.
48. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. The journals of gerontology Series A, Biological sciences and medical sciences 2001;**56**(3):M146-56
49. Buurman BM, van Munster BC, Korevaar JC, et al. Variability in measuring (instrumental) activities of daily living functioning and functional decline in hospitalized older medical patients: a systematic review. Journal of clinical epidemiology 2011;**64**(6):619-27.
50. Brane G, Gottfries CG, Winblad B. The Gottfries-Brane-Steen scale: validity, reliability and application in anti-dementia drug trials. Dementia and geriatric cognitive disorders 2001;**12**(1):1-14
51. Dencker K, Gottfries CG. Activities of daily living ratings of elderly people using Katz' ADL Index and the GBS-M scale. Scandinavian journal of caring sciences 1995;**9**(1):35-40
52. Katz S, Ford AB, Moskowitz RW, et al. Studies of Illness in the Aged. The Index of Adl: A Standardized Measure of Biological and Psychosocial Function. JAMA : the journal of the American Medical Association 1963;**185**:914-9

Figure Legends

Figure 1. Relative importance of predictors in the multivariate “model without ADL” and the “full model”. Interaction terms and non-linear effects have been incorporated in the variables. A higher χ^2-df value indicates a stronger association. Control = the grouping variable from the original study. BMI = body mass index, eGFR = estimated glomerular filtration rate.

Figure 2. Kaplan-Meier estimates from the updated full model including customary risk factors and ADL. The participants have been stratified into four equally sized groups by quartiles of risk.

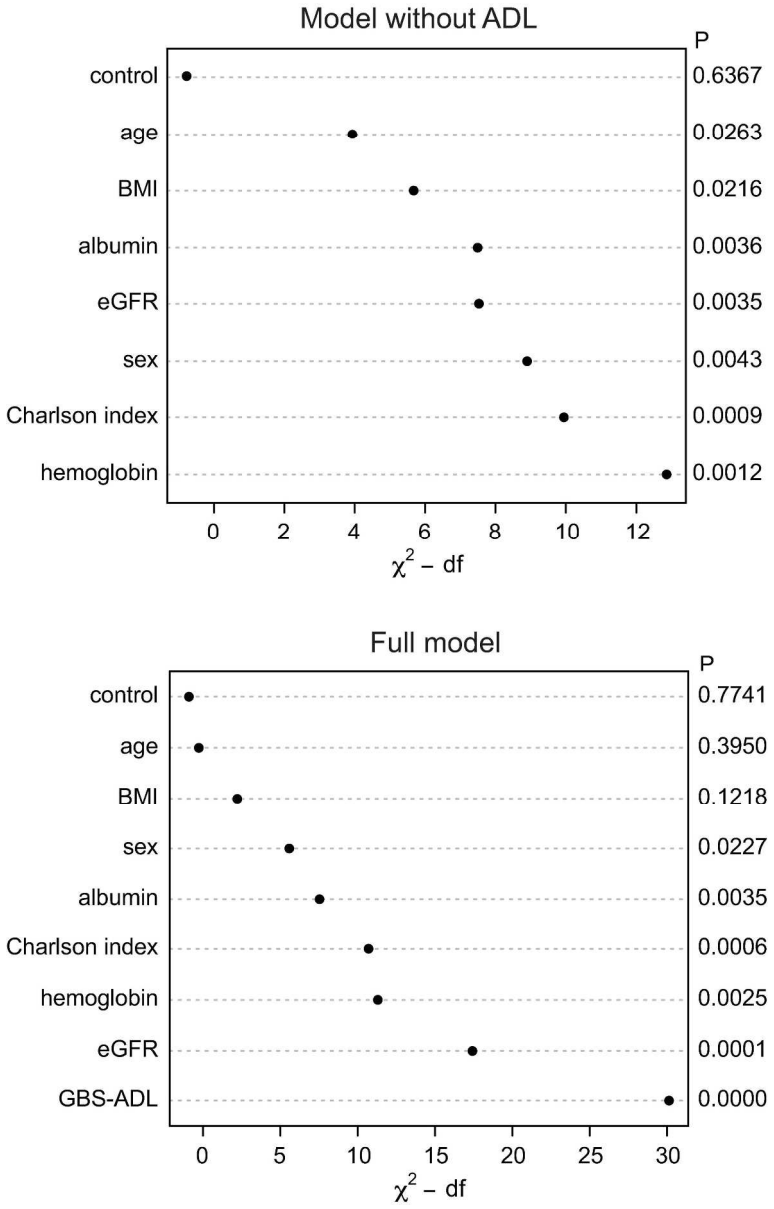

Figure 1. Relative importance of predictors in the multivariate "model without ADL" and the "full model". Interaction terms and non-linear effects have been incorporated in the variables. A higher χ2-df value indicates a stronger association. Control = the grouping variable from the original study. BMI = body mass index, eGFR = estimated glomerular filtration rate.

189x298mm (300 x 300 DPI)

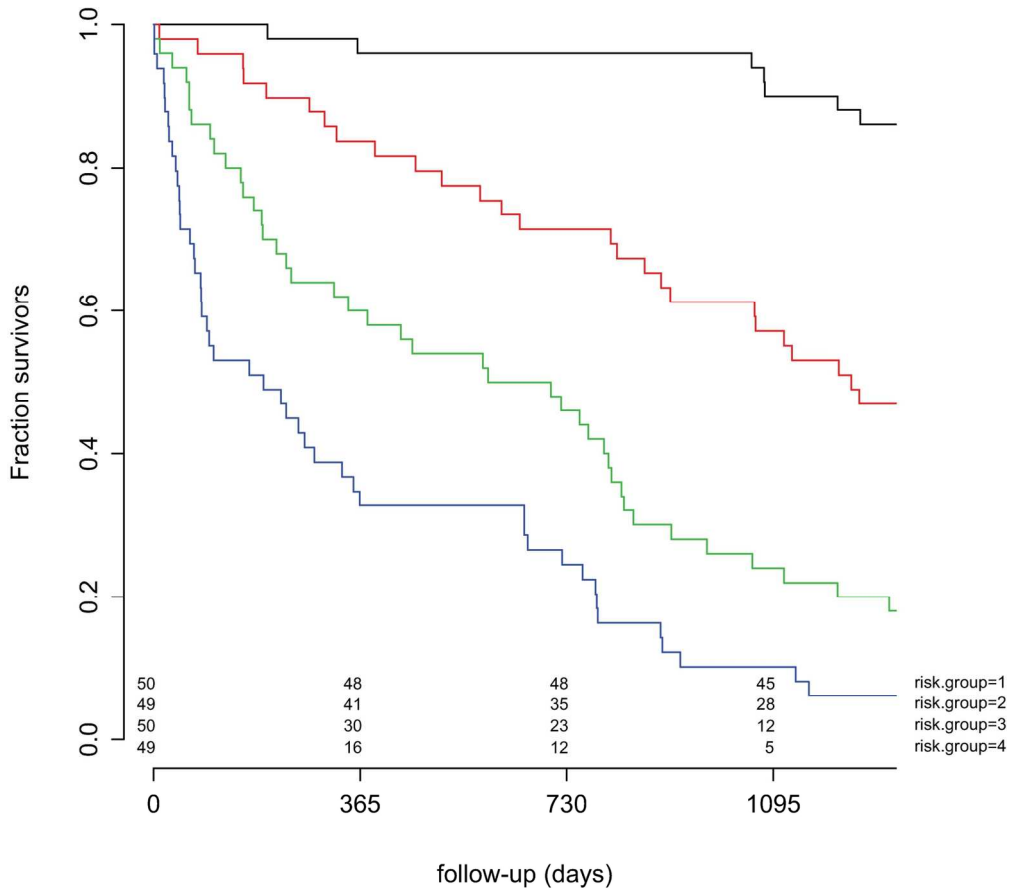

Figure 2. Kaplan-Meier estimates from the updated full model including customary risk factors and ADL. The participants have been stratified into four equally sized groups by quartiles of risk.

149x132mm (300 x 300 DPI)

Appendix 1 –statistics

General aspects

The overall aim was to compare ADL with the best possible model containing customary risk factors.
To achieve this, two models were developed and compared, the "model without ADL" and the "full model".
Data was originally stored in an SPSS file. All data analysis was performed in R [1]. Code is provided in appendix 2.

1. Outcome

Generally, it is important to describe the quantity of cases with missing outcome and to determine if there are any underlying patterns. Otherwise, simple exclusion may affect representativity [2].

In the study

Survival status was determined using the local region's electronic registry on the 6th February 2014. The time variable was defined as days from discharge to death or censoring at study endpoint, whichever came first. Those surviving at endpoint had been followed for a median of 1428 days (range 1312-1548). The baseline survival function is shown in figure e1.

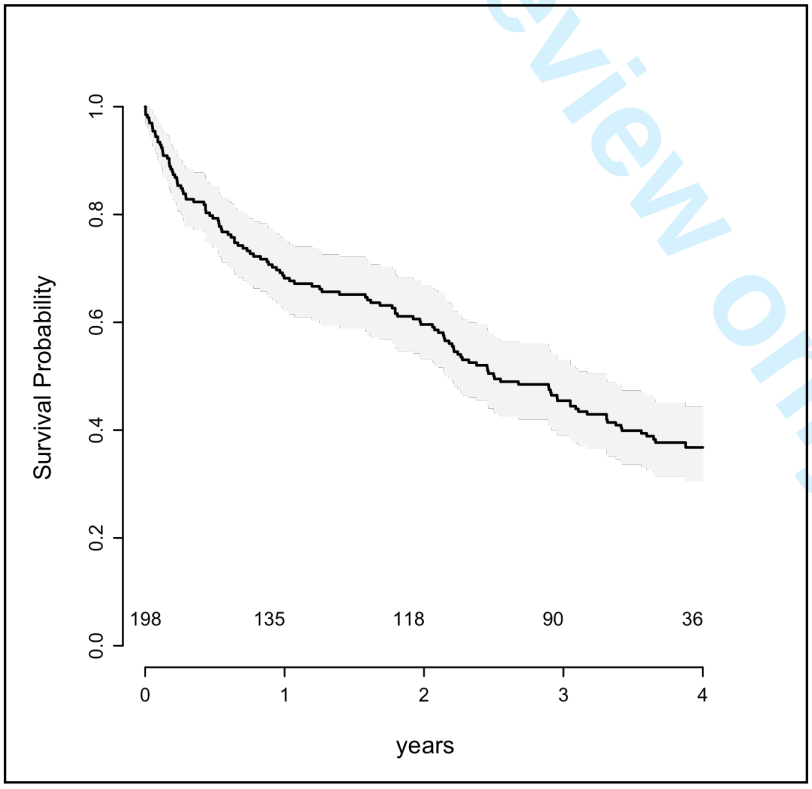

figure e1. Baseline survival function.

In our study, two cases were missing survival status due to having moved abroad (no longer in the region's registry) Hypothetically, these cases could be assumed to be in better health (severely diseased patients are unlikely to move abroad?). However, they were considered too few to affect representativity and were discarded from further analysis. Thus, the number of cases decreased from 200 to 198.

2. Crude analysis

Before any modifications are done to a variable, a crude analysis for the intented outcome could be of interest, to obtain an initial estimate of the effect of the predictor

In the study
Bivariate Cox proportional hazards regressions were carried out separately for all variables, including only outcome and the variable. All variables were treated in their original form, on their original scale. Observations with missing values were excluded from crude analysis. Data is presented with β coefficients, Standard errors, Wald χ^2, p value and hazard ratios in table 2 in the article.

In the crude analysis, all variables/potential predictors were statistically significant except sex, control/intervention status in the original study and the ADL item "food intake". Regarding the latter, the distribution was severely skewed, with only 18 cases (9%) having a non-zero value. To obtain a preliminary ranking of importance, the variables were sorted by decreasing Wald χ^2 in the table in the article.

In crude analysis, all separate GBS-ADL items were included but in further multivariate analysis, only the total GBS-ADL score was used, to avoid fitting too many variables and multicolinearity (the ADL items were intercorrelated at r = 0.8-0.9)

3. Missing data

In general, it is important to analyse missing data patterns in predictors. The first step is to determine the quantity of missing data. The second step is whether data is missing completely at random or if there are underlying patterns. When these prerequisites have been fulfilled, there are several approaches to missing data:

1. Listwise deletion, discarding all observations with any missing data points. The advantage of this approach is that no "manipulation" is done. Therefore, this method may seem intuitively most correct. The obvious disadvantage is that sample size could be substantially diminished. In addition, representativity could be affected, if missing a variable is systematically associated with other characteristics.

2. Using simple imputation. This technique substitutes missing values with the mean, mode or median value. This could be acceptable only if the variable is missing completely at random and the percentage of missing values small.

3. Using a more complex imputational technique. This approach uses customised regression models including all other covariates to obtain a stable prediction of the missing values. This method has been described and emphasized in several publications [3-7]. When using complex imputations, single or multiple imputations could be chosen. In the latter case, a separate

dataset is analysed for each imputational iteration, leading to a much larger complexity in the analysis.

In the study

When analysing the quantity of missing data, eGFR was missing in one case, BMI in three cases. Albumin was missing in 17 (9%) cases. BNP was missing in 113 (56%) observations.

The BNP variable was discarded from further analysis, as it had more than 50% missing. BMI and eGFR were considered missing completely at random. However, we found that cases with missing albumin were predominantly female (15 female vs. 2 male, $\chi^2 = 3.32$, $p = 0.056$) and had lower score on Charlson comorbidity index (1.47 vs. 2.33, $F = 11.3$, $p = 0.002$). Thus, excluding cases with missing albumin would affect representativity. Discarding the albumin variable would affect the overall aim, to compare ADL with the best possible traditional model. Therefore, the missing values in BMI, eGFR and Albumin were imputed using a single conditional imputation method (with the `transcan` function in R). In total, the effect of imputations was very small on the variable properties, as shown below.

Variable	β	S.E	Wald X2	p value	HR (95% CI)
Albumin, g/L, n = 181	-0.064	0.018	13.1	<0.001	0.94 (0.90 - 0.97)
- with imputation, transcan	-0.066	0.017	14.7	<0.001	0.94 (0.91 - 0.97)
eGFR, ml/min, n = 197	-0.029	0.005	29.3	<0.001	0.97 (0.96 - 0.98)
- with imputation, transcan	-0.029	0.005	29.2	<0.001	0.97 (0.96 - 0.98)
BMI, kg/m^2 , n = 195	-0.053	0.020	7.4	0.007	0.95 (0.91 - 0.99)
- with imputation, transcan	-0.053	0.020	7.4	0.006	0.95 (0.91 - 0.99)

Table e1. Effect of imputation on variable properties.

4. Variable considerations

Extreme outliers

In regression, outliers may be defined as observations with more than 3 interquartile ranges over the third quartile or below the 1st quartile. Such extreme values may affect a regression model significantly. First data entry errors should be considered and pursued. Then the biological plausibility should be considered. If plausible, we may consider a truncation at the 99th or 1st percentile [8].

In the study

In our study, data screening revealed, that for eGFR there was one extreme outlier with an estimated value of 198 ml /min (> 6 IQR over 3rd quartile), see boxplot.

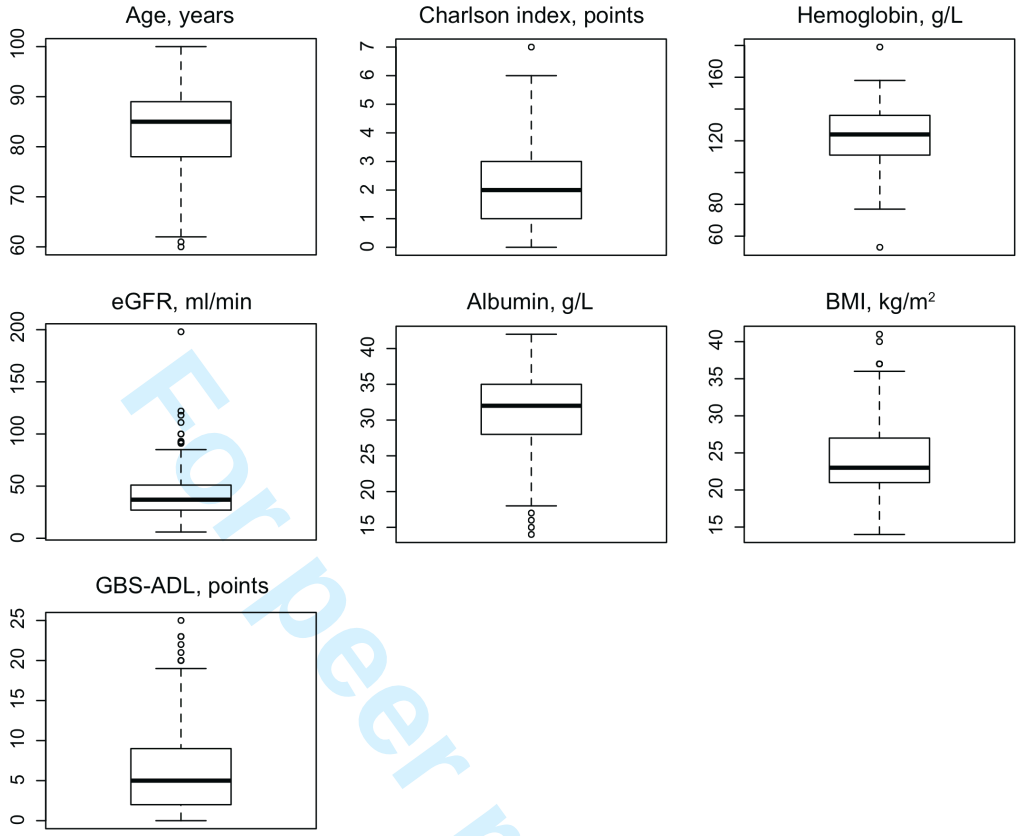

Figure e2. Boxplots of the continuous predictors. eGFR = Glomerular filtration rate, BMI = Body mass index, ADL = Activities of daily living.

This case was screened for data entry errors but none were found. Regarding biological plausibility, eGFR was measured with the Cockcroft-Gault formula ((140-age) * weight * constant)/Serum Creatinine in μmol/L, where the constant is 1.23 for men and 1.04 for women. Thus, GFR was not measured directly, but estimated and sensitive to extreme values in both serum creatinine, age and body weight. With this reservation, we considered the value to be biologically plausible. However, we did not consider it clinically important to compare one elderly patient with 198 ml/min in eGFR with another with 120 ml/min with regard to mortality. Therefore, eGFR was winsorized at the 99th percentile (118 ml/min). This led to a slightly improved fit in univariate performance.

Variable	β	S.E	Wald X2	p value	HR (95% CI)
eGFR, ml/min	-0.029	0.005	29.2	<0.001	0.97 (0.96 - 0.98)
- winsorized at 99th percentile	-0.029	0.005	29.8	<0.001	0.97 (0.96 - 0.98)

Table e2. effect of winsorization on variable properties.

Non-linearity

Most regression model assume that the predictors are linearly related to the outcome. However, non-linear relationships, such as U-shapes, for continuous variables are common.

There are several ways to address non-linearity:

First, assuming that the variable is linear. The advantage of this approach is that it results in an easily interpreted main effect, for example the Hazard Ratio in survival analysis. This is

the approach used in our crude comparisons. However, the approach is potentially problematic. For hemoglobin, this would mean that the risk difference between two individuals with 170 and 130 g/L would be the same as between two with 90 and 50, respectively. In addition, this approach cannot handle U-shaped risks, it is likely that someone with 200 g/L in Haemoglobin with dehydration or polycytemia does not have better survival than someone with 140 g/L

Second, to dichotomize the variable, using a previously established cut-off, is another frequently used approach. However it is not recommended as it ignores a lot of information [9]. In our example, applying the WHO cut-off for anemia (120 g/L for women, 130 for men) would attribute the same risk for an individual with Hb of 119 g/L as for one with 53 g/L (the lowest in our material).

Third, to categorise the variable into categories that are clinically important, creating dummy variables. This approach could handle U-shaped risks. However, previously defined clinically important categories are needed and several degrees of freedom is spent in the analysis. As with a dichotomous transformation, all cases within a category are attributed the same risk.

Fourth, to use a more complex fitting function, such as a restricted cubic spline [10, 11]. This approach uses so called knots, point estimates where the risk is determined. A cubic function is used to fit the function between knots. Near the ends the risk is modelled linear.

In the study

We prespecified Hemoglobin to be non-linear and tried the approaches above, see figure e4. We decided to use the 4-knot restricted cubic spline as both the best performance and was most appropriate from a clinical perspective. The knots were placed at the 5th, 35th, 65th and 95th percentiles where Hb was 92.25, 118, 130 and 148.15, respectively. The resulting function to fit Hb was:

```
2.2712251-0.017758194* hb-6.2295666e-06*pmax(hb-
92.25,0)^3+5.8240197e-05*pmax(hb-118,0)^3-7.7559735e-
05*pmax(hb-130,0)^3+2.5549104e-05*pmax(hb-148.15,0)^3
```

As opposed to the easily interpreted hazard ratio from the linear function, this is not easy to interpret without a graph, the graphic display of the four approaches is presented in figure e3.

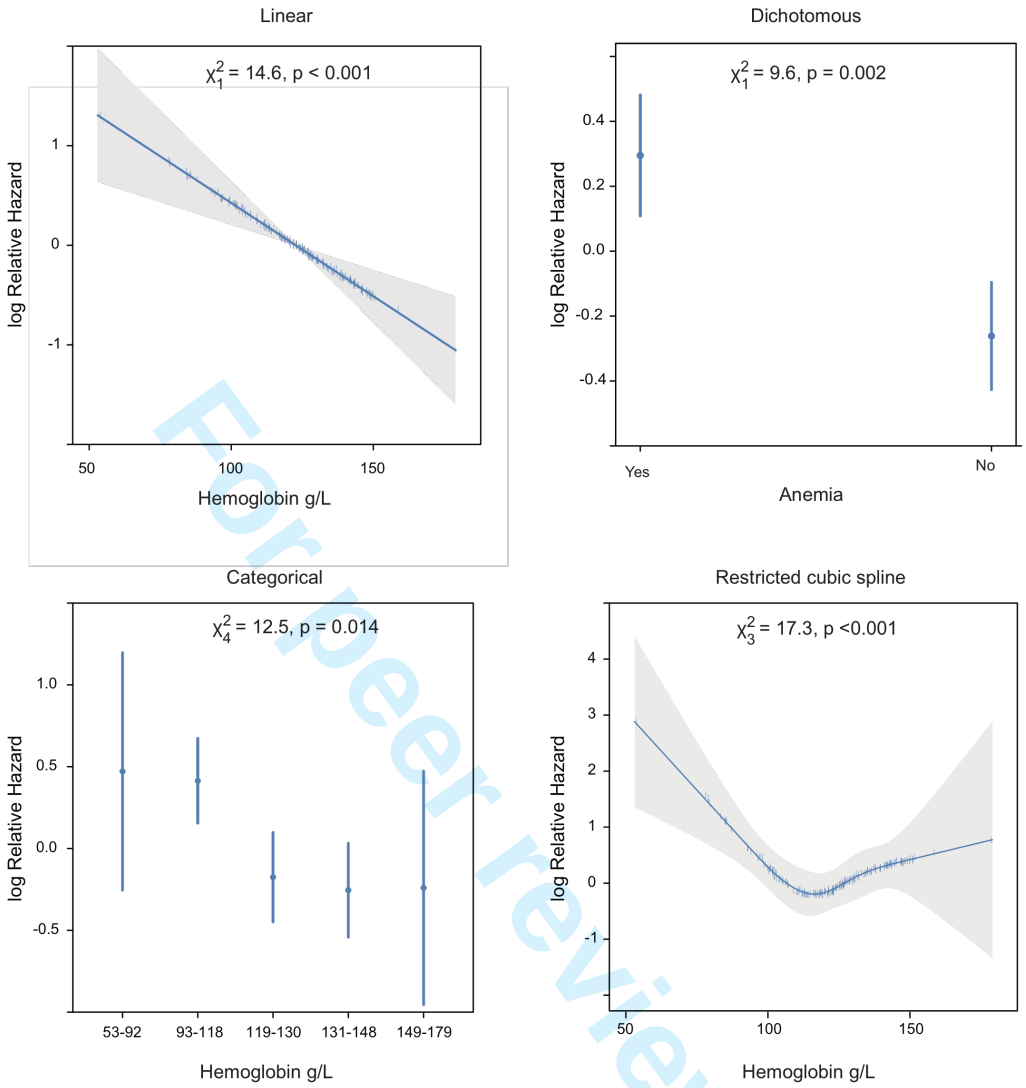

Figure e3. Different transformations of Hemoglobin. For dichotomous, the WHO definition of anemia is used. For categorical, the 5th, 35th, 65 and 95th percentiles were used, for easier comparison with the spline fit.

Apart from Hemoglogin, all other variables were bivariately tested for non-linearity by using 4-knot splines followed by ANOVA tests to determine if there was a significant non-linear component. GBS-ADL showed significant non-linearity and different codings were tested. We tested dichotomizing at the median and categorizing at the quartiles. A polynomial showed good fit but was not clinically plausible, with decreasing risks at the higher end of ADL impairment. The restricted cubic spline resembled a log fit and indeed the log fit was chosen, with fewer degrees of freedom spent, see figure e4. No other variables showed significant non-linear effects.

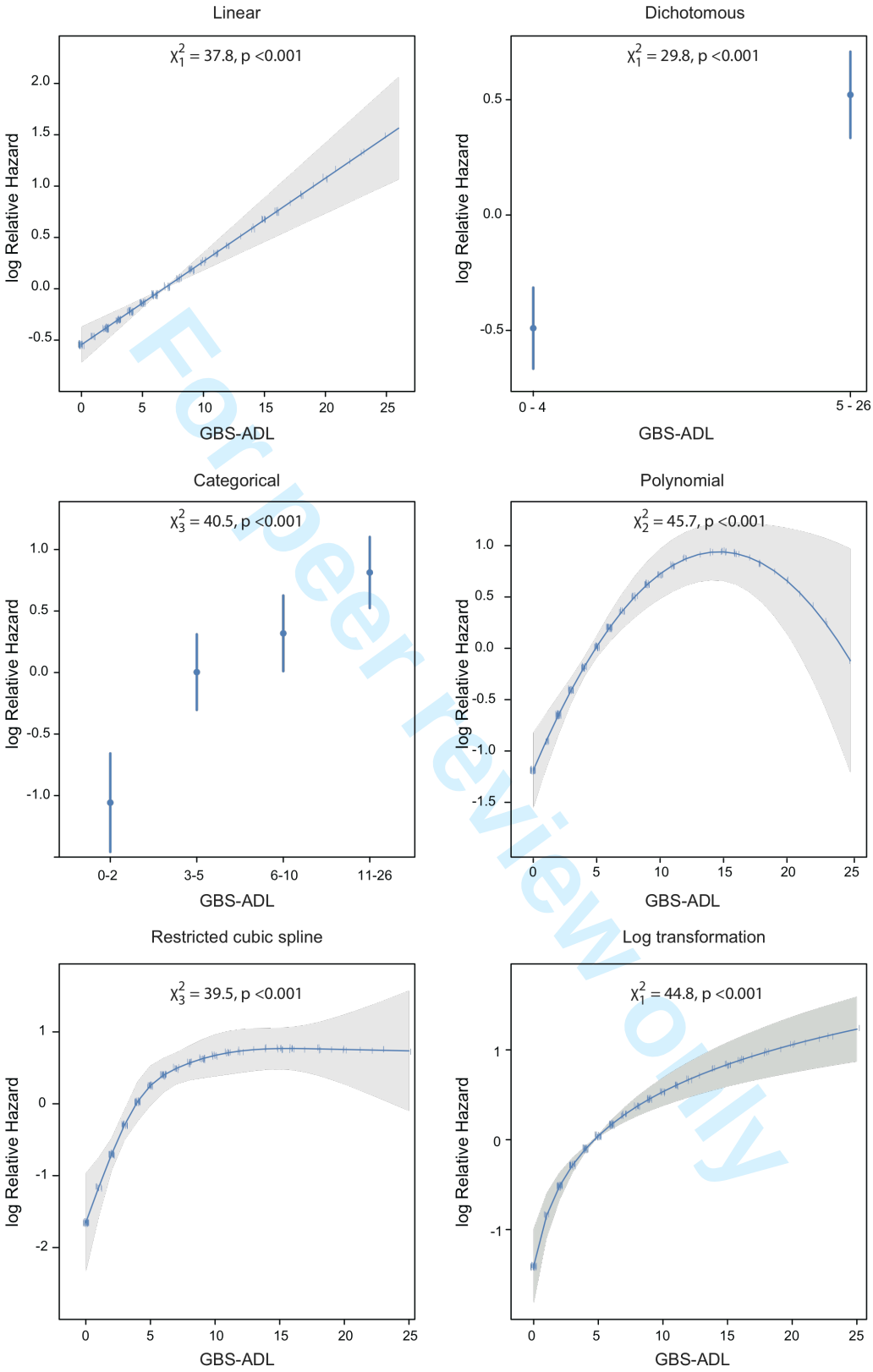

Figure e4. Different transformations tested for GBS-ADL.

5. Fitting the multivariate models

In the study

The two models were fitted, using the imputations and transformations above. The "model without ADL" used the covariates age, sex, charlson comorbidity index, albumin, BMI, eGFR, control/intervention status, and hemoglobin fitted as a restricted cubic spline The "full model" also included log(GBS-ADL).

6. Multivariate Diagnostics - Multicolinearity

Predictors with strong intercorrelations could cause interpreting problems, this is tested using the variance inflation factor (VIF). The interpretation of VIF has been disputed, a rule of thumb saying that VIF > 4 or > 10 signals a problematic multicolinearity problem have been suggested. However, these cut-offs may be too low, as a VIF over 10 could be acceptable [12]. To address multicolinearity, clustering of variables or data reduction could be applied.

In the study

In our models, all variables were simultaneously tested for colinearity. VIF Values were ranging between 1.02 and 1.47 in the "model without ADL" and between 1.10 and 1.52 in the "full model". The strongest bivariate correlation was between age and eGFR (r = -0.49). Thus, no apparent multicolinearity was present and no further action was taken.

7. Interactions – additivity assumption

A two-way interaction occurs when the effect of one predictor is dependending on the value of one other predictor. There are several recommendations regarding the number of interactions to test for. Only clinically plausible interactions could be tested, however, this requires prior knowledge. Another strategy is to test for all possible interactions, this requires a very large sample, to avoid overfitting. A compromise is to do a pooled interaction test for each variable and if the test is significant, the specific interactions are pursued [11].

In the study

We did not have prespecified interactions for ADL and the sample size did not permit testing for all possible interactions. Therefore we opted for a global test approach. As we did not want to give ADL any advantages compared to the other variables, we also performed global tests for the other variables, one at a time. In the "model without ADL", the global test was significant for sex and BMI and an interaction term of sex * BMI was found (low BMI was a risk factor in men, not in women). This interaction was included in the model. In the "full model" another interaction, GBS-ADL*eGFR, was also found (the effect of impaired GBS-ADL was higher when eGFR was less imparied and vice versa). One interpretation of this interaction could be that impaired GBS-ADL is associated with weight loss and thus lower eGFR. To test properly for this we would need to apply three-way interactions (such as GBS-ADL*BMI*eGFR), which was beyond the scope of this paper.

8. Assumption of proportional hazards

The assumption of proportional hazards is the assumption that hazards from predictors do not vary over time. Proportional hazards can be tested in several different ways. Graphically, schoenfeld residuals are often plottet against time, then a straight line at zero is ideal. There are also different approaches to compensate for non-proportional hazards, the most common being adding an interaction term with time.

In the study

The proportional hazards assumption was first tested using a global test (`cox.zph` in R) as well as specific tests for all variables. In the "model without ADL", the global test gave a p value of 0.72 and in the "full model" a p value of 0.70, signalling no violations of the PH assumption. The variable closest was eGFR, with a p value of 0.14. For eGFR, a schoenfeld residual plot is shown in figure e5. No further action was taken.

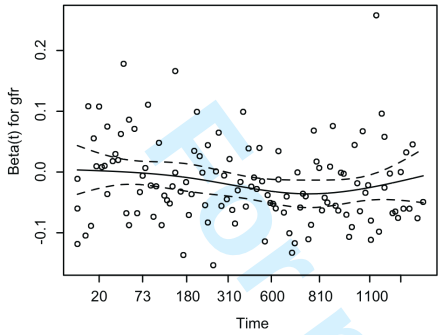

figure e5. Schoenfeld residual plot for eGFR.

9. Influential observations

With small sample size, a few influential observations could affect a model significantly. One way to screen for influential observations is by using what is called dfBeta, that shows to what extent the regression coefficient would change, if that case should be removed. Every case is designated a dfBeta value for each variable. Often, standardised dfBetas, with a cutoff of 0.20 is used to signify an influential observation. Thus, if deleting one observation led to a change in a predictor's β coefficient of more than 0.2 standard error, that observation was noted. For variables of specific interest, a sensitivity analysis could be performed without the observations with dfBeta > 0.2 to determine whether the effect is mainly due to a few highly influential observations.

In the study

In the "model without ADL", a total of 23 (12%) observations had any DfBeta > ±0.20. The lowest dfBeta was -0.39 and the highest 0.32. In the "full model", 21 observations were considered influential. DfBetas ranged from -0.46 to 0.48. Nine cases had a dfBeta > ± 0.20 for GBS-ADL and/or its interaction with eGFR. A sensitivity analysis was done, with these nine observations excluded. In that model the overall χ^2 increased from 123 to 124 and the GBS-ADL χ^2 from 32 to 37. Thus, the effects of GBS-ADL in the "full model" were not due to a few influential observations. In all further analysis the influential observations were kept in the model.

10. Relative contribution of variables

Describing the main effects of predictors including non-linear terms and interaction terms is not as intuitive as for simpler models, using Hazard Ratios. This is especially true if the model contains continuous-by-continuous interactions.

In the study

To obtain an estimate of the relative importance of the different predictors, we used the anova approach, developed by Harrell (`anova.rms` in R)[11]. Simple anova plots were included in

the article as figure 1. Plots of the variable effects are shown below in figure e6. In these plots, interaction terms have been incorporated into the variables' relative importance.

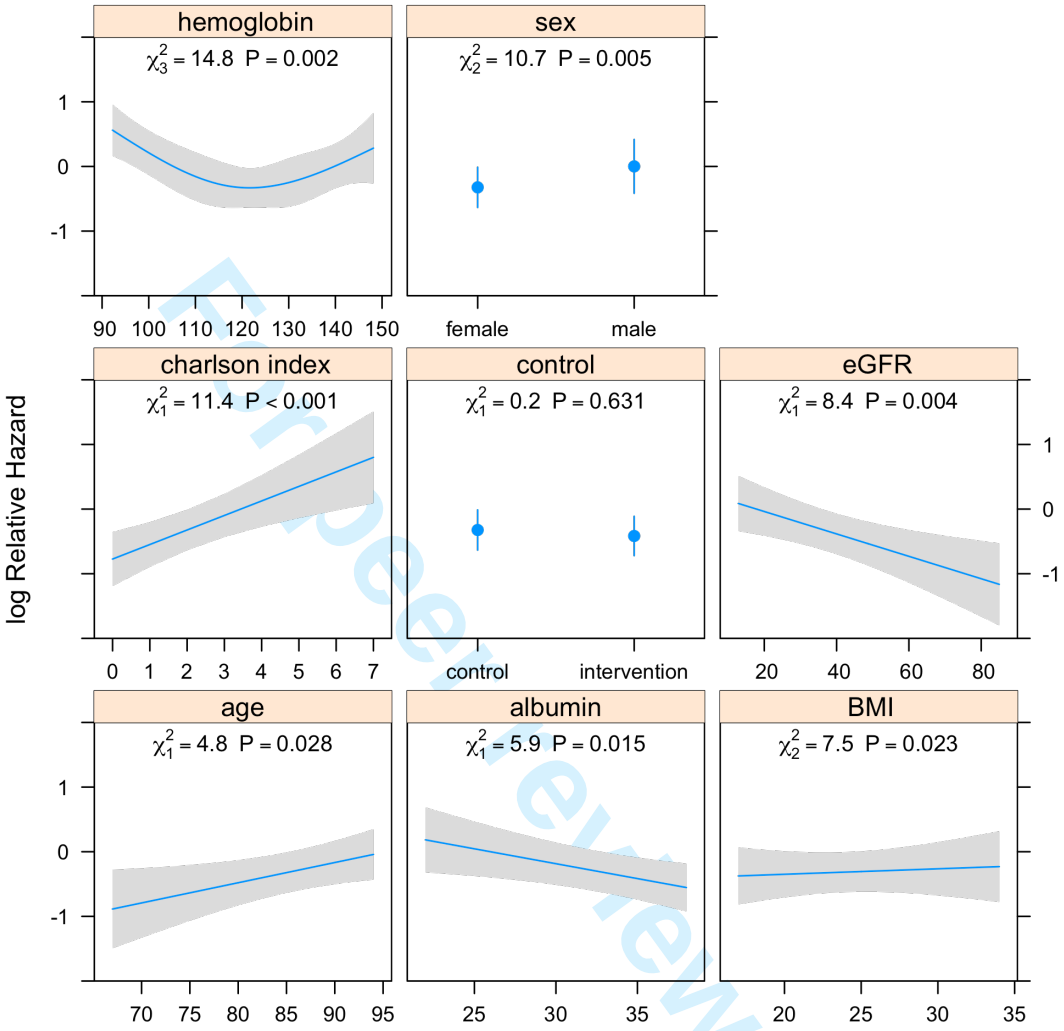

Figure e6a. Plot of variable effects in the "model without ADL"

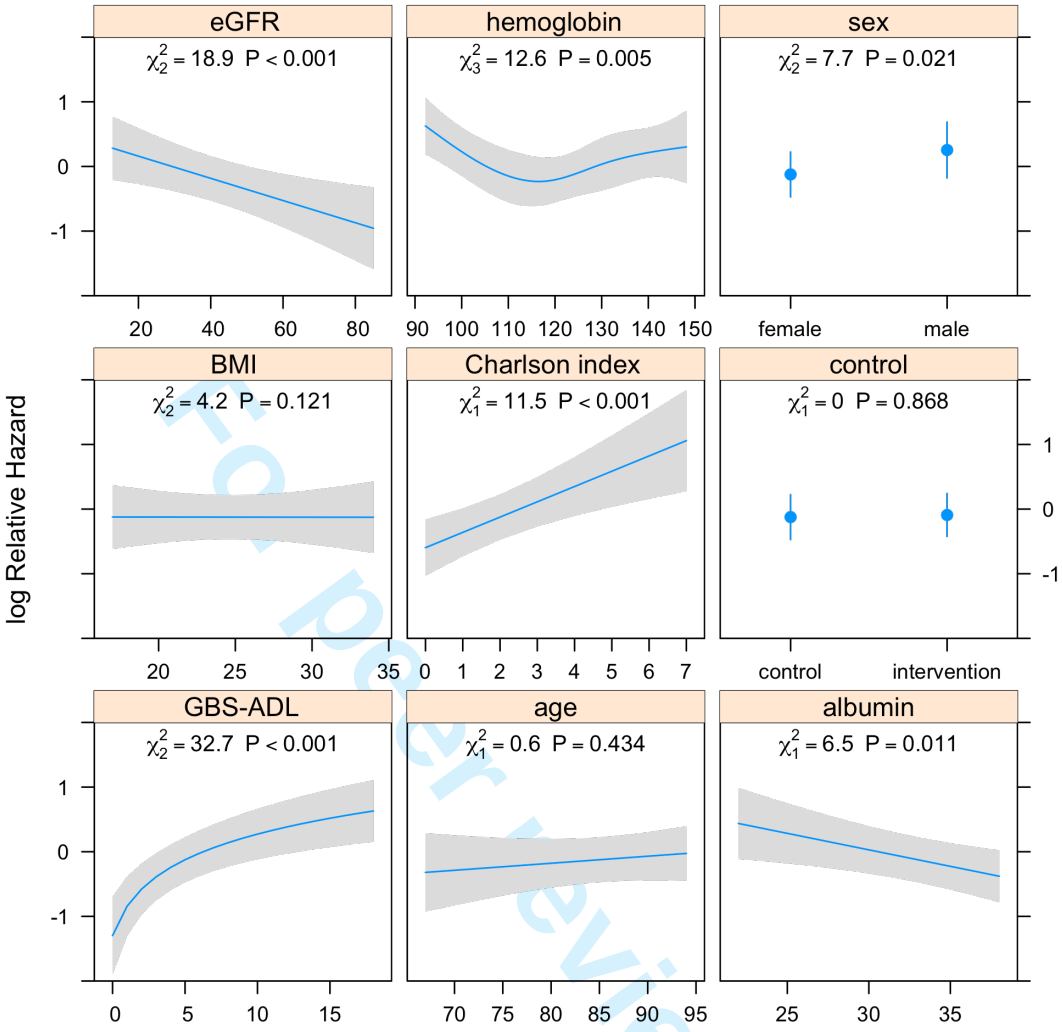

Figure e6b. Plots of the variable effects in the "full model".

11. Added value of an added variable

There are several ways to determine the added value of a variable in a regression model.

a. Likelihood ratio test. With two nested models (where the smaller model is also a part of the full model) a Likelihood ratio test could be performed as a χ^2 test over df = number of additional independent variables in the new model.

In the study

The results are shown in table 3 in the article. For the "model without ADL", LR χ^2 was 78.4 and for the "full model" 121.0. The degrees of freedom were 11 and 13, respectively. Therefore the LR test resulted in a χ^2 (df = 2, N = 198) = 42.5, p < 0.001. Thus the "full model" had a significantly better fit.

b. Discrimination, measured with the C, or concordance, statistic. The C statistic is the probability that, in a case-control pair, the case (deceased) will be given a higher predicted risk from the model than the control (survivor). C statistics range from 0.5 (coin toss, useless) to 1.0 (perfect discrimination). In logistic regression (without time-to-event data), the c

statistic is the same as ROC. For survival analysis, time is incorporated, so a case at time t is compared with a survivor at time t, albeit this survivor could be dead at time t+1 (the next day). C statistics in survival analysis are often lower than ROC in logistic analysis. In addition, there are several different ways to calculate c statistic for time-to-event data.

In the study
We chose the method by Uno, to be able to compare between models. The "model without ADL" had a c statistic of 0.72 and the "full model" of 0.78. We set the follow-up time to 1428 days, as this was our median follow-up time of survivors. C statistics from the two models were compared using the method described by Uno et al. in the SurvC1 package [13]. Difference in c statistic between the model without ADL and the full model was 0.058 (95% CI = 0.022 - 0.094, p value 0.002).

c. NRI >0. Continuous net reclassification index[14, 15]. This index determines to what extent adding a new variable to a model leads to a change in the correct direction in predicted risk for each observation at time t (towards higher risk for deceased, towards lower for survivors). NRI>0 ranges from 0 (no increased value, useless) to 1(all observations reclassified in the right direction). NRI>0 has been shown to be more sensitive than change in C index, especially when the baseline model has a good performance. NRI>0 only describes the share of observations that have been reclassified, it does not quantify the amount of change in risk. Thus, it cannot distinguish between adding a variable that increases the predicted mortality risk for all cases with 1% or one that increases it with 50%.

For interpretation, the original NRI > 0 has been compared to the effect size of the added variable, where NRI>0 of 0.6 should be considered strong, 0.4 intermediate and below 0.2 weak [16]. However, after the initial development, Pencina et al. have suggested that ½ NRI>0 shoud be reported, as an average[15]. This is also what is given by the `IDI.INF` function in the `SurvIDINRI` package in R.

In the study
In our study ½ NRI>0 (95%CI) was 0.42 (0.22-0.58) with a p value <0.001. Again the follow-up time was set to 1428 days, to avoid extensive censoring. By doubling the point estimate of ½ NRI>0, the original NRI>0 would be 0.84, indicating a substantial effect size of adding ADL.

IDI. Integrated discrimination improvement. Originally developed by Pencina et al. for logistic models, IDI has been extended to time-to-event data [14, 17]. While NRI>0 displays the percentage of observtations being reclassified in the desirable direction, IDI is related to mean change in predicted probabilities within cases and controls. IDI is similar to testing the difference in R2, or discrimination slope, in logistic regrssion. IDI and NRI with confidence intervals were calculated (using the `IDI.INF` function) with the method by Uno et al. [18].

In the study
IDI was 0.15 (95%CI 0.07-0.27, p < 0.001), indicating that the mean change in the correct direction was 15% (cases (deceased) were given 15% higher mortality risk by adding ADL while controls (survivors) were given 15% lower).

12. Overfitting - internal validation

Overfitting occurs when sample size is small. Then the fitted model becomes too optimistic and dependent on the present dataset. Thus, the findings will neither be reproducable nor

valid in other populations. Ideally a model could be tested in another population at another location and setting, what is known as external validation. If that is not possible, there are several ways to accomplish internal validation. The recommended approach is via bootstrapping. In bootstrapping a new dataset, of the same size as the original, is constructed from the original dataset by resampling with replacement (an observation could be selected several times). This dataset is then used to develop the model, which is then tested on the full original population. The difference in apparent performance and resampled performance is called optimism. The procedure is repeated 200-1000 times and the mean optimism is subtracted from the apparent performance estimates. This way an optimism-corrected estimate is obtained. In a future external population, this corrected estimate should be considered the best estimate possible[2, 8, 11, 19, 20].

In the study

Our aim was not to develop a valid prediction model, as this would have called for a larger sample size. Instead, we aimed to determine the relative imporance and added value of ADL when compared to the best possible traditional model. In the trade-off between overfitting and a well-performing traditional model, we empahsised the latter. A heuristic estimate of shrinkage would be χ^2 - d.f. / χ^2 = 123 - 14/123 = 0.89 for the "full model". However, this d.f. is falsely low as we tested many more interactions and transformations.

To better determine, the extent of overfitting, we carried out an internal validation with 1000 bootstraps. For the "model without ADL" and the "full model", the calibration slopes were 0.84 and 0.83, respectively, indicating a substantial amount of overfitting. The optimism-corrected R^2 was 0.26 vs. 0.39. Optimism-corrected c statistics were 0.69 and 0.76.

13. Updating the model - final nomogram and Kaplan-Meier curves

An overfitted model could be updated using a model that shrinks the regression coefficients. One such method is the LASSO (least absolute shrinkage and selection operator) model [21, 22]. LASSO could be used to both shrink factors as well as to eliminate variables.

In the study

Even if our aim was not to develop a valid prediction model, the amount of overfitting suggested that we should try to update the "full model". In a LASSO model, the interaction terms and non-linear terms were combined into single terms. We used the `coxpath` function in R. We considered the model with the lowest AIC (Akaike Information Criterion). In this model, the variable "control" was shrunk to zero, all other variables remained. The mean shrinkage was 0.84. The lasso path and AIC is shown in figure e7.

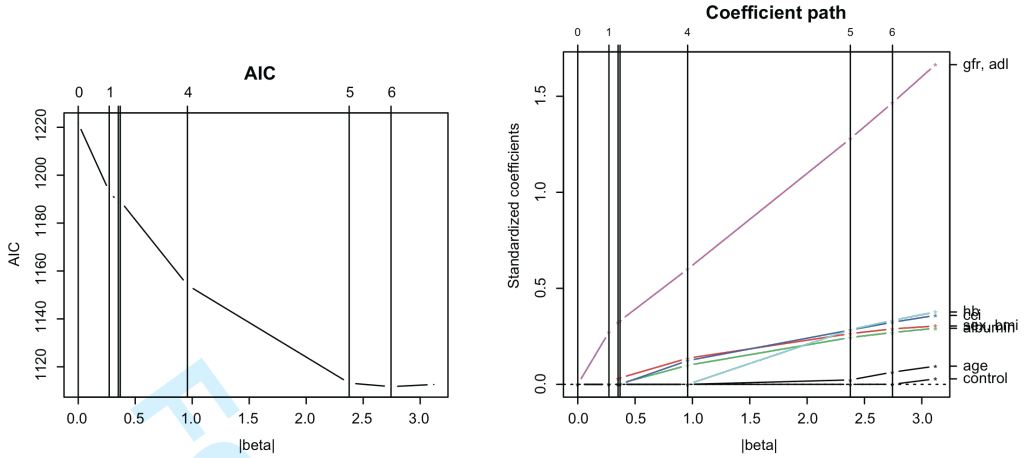

Figure e7. Lasso plots. AIC (Akaike information criterion) is lowest when control is set to zero.

A final nomogram, using the shrunk Lasso coefficients, was built, see figure e8. The cases were divided into four equally sized risk groups, by the quartiles of the linear predictor. To display the discrimination of the model, a kaplan-Meier curve was built, included in the article as figure 2.

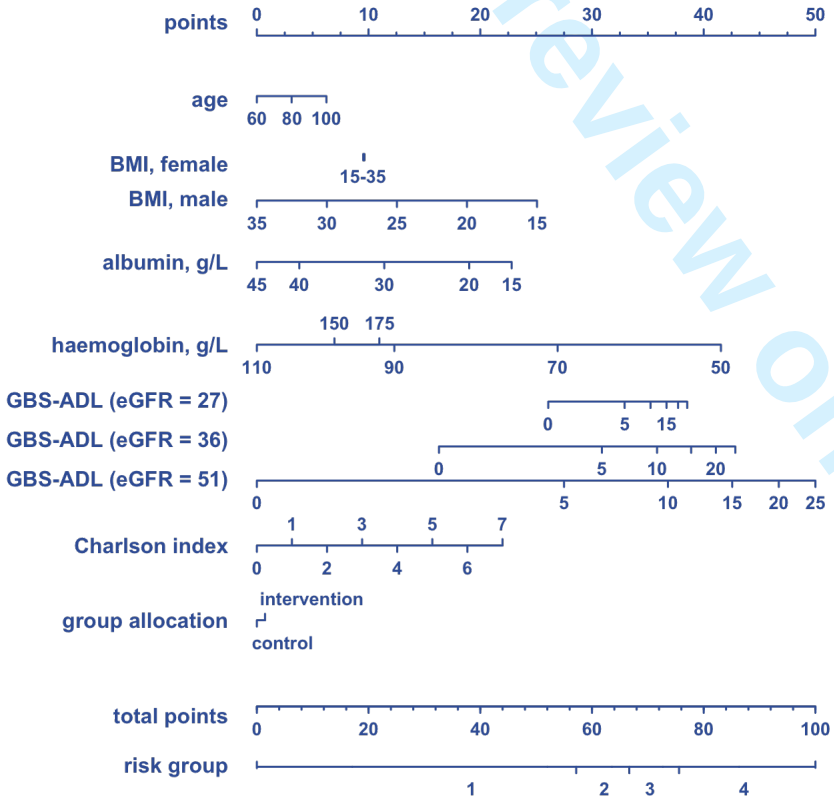

Figure e8a. Nomogram. Interpretation: For an individual, the variables are compared with the upper ”points” line, one at a time. These scores are then added for a total score that is plotted at the ”total points” line at the bottom. This could then be used to designate the person to a ”risk group” Notice the effect of interactions, low BMI is only a risk factor in men and the risk of GBS-ADL is moderated by eGFR, which is presented by median and quartiles. The cutoffs

in the nomogram for the risk groups are completely arbitrary here, created to obtain 4 equally sized groups. In another scenario, cutoffs could be established to obtain for example a group with 90% chance of 3-year survival.

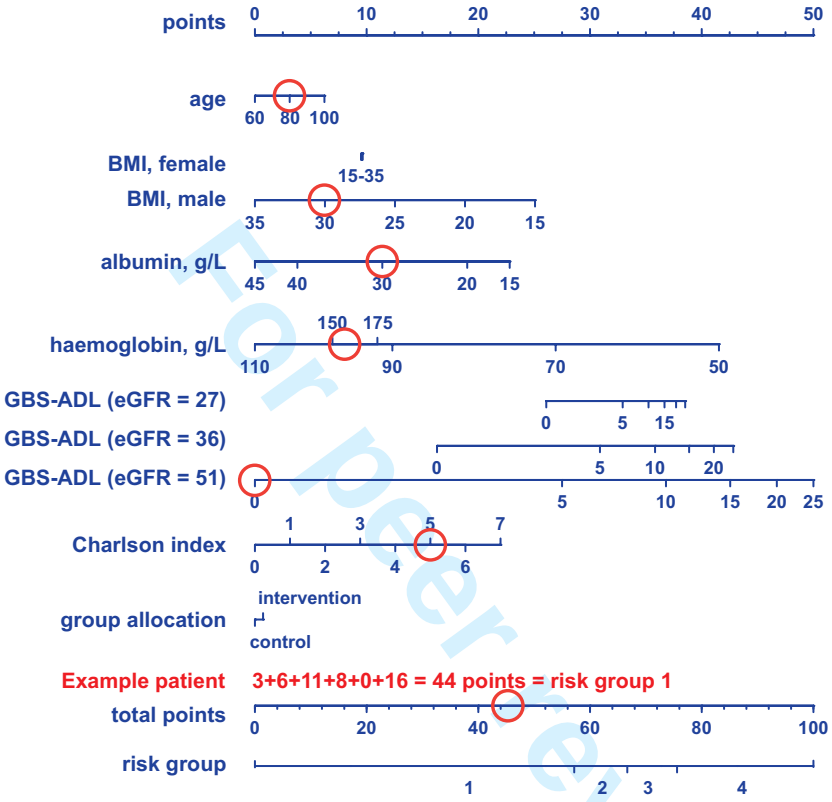

Figure e8b. Example of a scoring: This patient is 80 years old (3 points), male with BMI 30 (6 points), has an albumin of 30 (11 points) and a hemoglobin of 98 (8 points), normal kidney function and GBS-ADL (0 points) and a Charlson index score of 5 (16 points). The total score would be 3+6+11+8+0+16 = 44 points, placing this patient well within risk group 1. If this patient had all other variables constant but a functional decline, with a GBS-ADL score of 7, this would result in a total score of 44+30 = 74, placing the patient in risk group 3. The risk attributed to the functional decline would be equivalent to a hemoglobin drop from 98 to 55 g/L. Would it infer the same sense of urgency to the clinician?

References - statistical appendix

1. *R: A language and environment for statistical computing.* [computer program]. Version. VIenna; 2014.
2. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* Feb 28 1996;15(4):361-387.
3. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol.* Oct 2006;59(10):1087-1091.
4. Horton NJ, Kleinman KP. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *Am Stat.* Feb 2007;61(1):79-90.
5. Janssen KJ, Donders AR, Harrell FE, Jr., et al. Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol.* Jul 2010;63(7):721-727.
6. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Med Res Methodol.* 2010;10:7.
7. Moons KG, Donders RA, Stijnen T, Harrell FE, Jr. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol.* Oct 2006;59(10):1092-1101.
8. Steyerberg EW. *Clinical Prediction Models - A Practical Approach to Development, Validation, and Updating*. New York: Springer; 2009.
9. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med.* Jan 15 2006;25(1):127-141.
10. Harrell FE, Jr., Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst.* Oct 5 1988;80(15):1198-1202.
11. Harrell Jr FE. *Regression Modeling Strategies - With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer-Verlag; 2001.
12. O'Brien RM. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity* 2007-10-01 2007;41(5):673-690.
13. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med.* May 10 2011;30(10):1105-1117.
14. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med.* Jan 30 2008;27(2):157-172; discussion 207-112.
15. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med.* Jan 15 2011;30(1):11-21.
16. Pencina MJ, D'Agostino RB, Pencina KM, Janssens AC, Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol.* Sep 15 2012;176(6):473-481.

17. Chambless LE, Cummiskey CP, Cui G. Several methods to assess improvement in risk prediction models: extension to survival analysis. *Stat Med.* Jan 15 2011;30(1):22-38.

18. Uno H, Tian L, Cai T, Kohane IS, Wei LJ. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Stat Med.* Jun 30 2013;32(14):2430-2442.

19. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med.* Aug 30 2004;23(16):2567-2586.

20. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* Aug 2001;54(8):774-781.

21. Tibshirani R. Regression shrinkage and selection via the Lasso. *J. Royal. Statist. Soc B.* 1996;58(1):267-288.

22. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med.* Feb 28 1997;16(4):385-395.

```
##########################################################
##                                                      ##
##            Appendix 2 - R code                       ##
##                                                      ##
##########################################################

## Data import from SPSS   ##

> library(rms)
> library(Hmisc)
> data <- spss.get() #file path in brackets

## Basic properties, table 1 (article) ##

> names(data)

"Nr"          "age"         "sex"         "status"
"time"        "cci"         "dressing"    "eating"
"physical"    "spontaneous" "hygiene"     "toilet"
"adl"         "hb"          "gfr"         "albumin"
"bnp"         "bmi"         "control"

> describe(data)

##########################################################
## 1. outcome                                           ##
##                                                      ##
##########################################################

## Determining follow-up and censoring ##

> library (survival)
> S <- Surv(time, status)
> cens.time <- ifelse(status == ”alive”, time, NA)
summary(cens.time)

## Baseline survival plot, figure 1 (appendix 1) ##

> S.years <- Surv(time/365.25, status)

> survplot(npsurv(S.years~1), xlab = "years", xlim =
c(0,4), time.inc = 1, lwd = 1.5, n.risk = T, y.n.risk =
0.05, cex.n.risk = 1, adj.n.risk = 0.5)

## discarding cases with missing outcome ##

> missing.outcome <- is.na(data$status)
> data <- data[missing.outcome == FALSE,]
> attach(data)

##########################################################
## 2. Crude analysis                                    ##
##                                                      ##
##########################################################
```

```
## Crude analysis for all predictors, table 2 (article)##

> summary(coxph(S~age))
> summary(coxph(S~sex))
> summary(coxph(S~cci))
> summary(coxph(S~dressing))
> summary(coxph(S~eating))
> summary(coxph(S~physical))
> summary(coxph(S~spontaneous))
> summary(coxph(S~hygiene))
> summary(coxph(S~toilet))
> summary(coxph(S~adl))
> summary(coxph(S~hb))
> summary(coxph(S~gfr))
> summary(coxph(S~albumin))
> summary(coxph(S~bnp))
> summary(coxph(S~bmi))
> summary(coxph(S~control))

## Discarding the separate ADL items ##

> data <- data[,-c(7:12)]

############################################################
## 3. Missing data                                        ##
##                                                        ##
############################################################

## Defining covariates ##

> covs <- data[, c("age", "sex", "cci", "adl", "hb",
"clearance", "albumin", "bnp", "bmi", ”control”)]

## Plotting missing, figure 2 (appendix 1) ##

> missing <- naclus(covs)
> naplot(missing, which = ”na per var”)

## Discarding the BNP variable ##

> data <- data[,-8]

## Determining associations with missing albumin ##

> missing.albumin <- ifelse(is.na(albumin), 1, 0)
> lrm(missing.albumin~age+sex+cci+clearance+bmi+adl+hb)
> chisq.test(missing.albumin, sex)
> oneway.test(cci~missing.albumin)
> tapply (cci, missing.albumin, mean)

## Creating a transcan object for imputation ##

> trans <-transcan(~age + sex + cci + adl + hb +
clearance + albumin + bmi + control, imputed = T, data =
data)
```

```
## Imputing albumin, gfr and bmi ##

> albumin.imputed <- as.integer(impute(trans, albumin,
data = data))
> gfr.imputed <- as.integer(impute(trans, gfr, data =
data))
> bmi.imputed <- as.integer(impute(trans, bmi, data =
data))

## Testing imputed variables, table 1 (appendix 1) ##

> summary(coxph(S~albumin.imputed)
> summary(coxph(S~gfr.imputed)
> summary(coxph(S~bmi.imputed)

## Imputed variables put into original dataset ##

> data$gfr <- gfr.imputed
> data$albumin <- albumin.imputed
> data$bmi <- bmi.1
> attach(data)

############################################################
## 4. Variable considerations                             ##
##                                                        ##
############################################################

## Screening for outliers, figure 3 (appendix 1) ##

> par(mfrow = c(3,3))
> boxplot(age)
> boxplot(cci)
> boxplot(hb)
> boxplot(gfr)
> boxplot(albumin)
> boxplot(bmi)
> boxplot(adl)

## Winsorising at 99th percentile, table 2(appendix 1) ##

> describe(gfr)
> gfr.winsorised <- ifelse(gfr > 118, 118, gfr)
> summary(coxph(S~gfr.winsorised))
> data$gfr <- gfr.winsorised
> attach(data)

############################################################
## Transformations for haemoglobin figure 4, appendix 1##
############################################################

## Range is obtained ##

> describe(hb)

## Linear model fitted and plotted ##
```

```
> hbfit.linear <- cph(S~hb, data = data)
> plot(Predict(hbfit.linear, hb = seq(53, 179, by = 1)),
xlab = "Haemoglobin g/L" , anova = anova(hbfit.linear),
pval = T, data = llist(hb))

## Dichotomous model##

> data$anemia <- ifelse(sex == ”male”, ifelse(hb < 130,
”yes”, ”no”), ifelse (hb < 120, ”yes”, ”no”))
> dd <- datadist (data)
> options(datadist = ”dd”)
> hbfit.dichotomous <- cph(S~anemia, data = data)
> plot(Predict(hbfit.dichotomous), xlab = "Anemia", anova
= anova(hbfit.dichotomous), pval = T)

## Categorical model ##

> data$hbcat <- ifelse(hb < 92.25, 1, ifelse(hb < 118, 2,
ifelse(hb < 130, 3, ifelse (hb < 148, 4, 5))))
> dd <- datadist (data)
> options(datadist = ”dd”)
> hbfit.categorical <- cph(S~as.factor(hbcat), data =
data)
> plot(Predict(hbfit.categorical), anova =
anova(hbfit.categorical), xlab = ”Haemoglobin g/L”, pval
= T)

## Restricted cubic spline ##

> hbfit.spline <- cph(S~rcs(hb,4), data = data)
> plot(Predict(hbfit.spline, hb = seq(53,179, by = 1)),
anova = anova(hbfit.spline), pval = T, xlab =
"Haemoglobin g/L", data = llist(hb))

## Testing other continuous variables ##

> agefit <- cph(S~rcs(age,4), data = data)
> anova(agefit)

> ccifit <- cph(S~rcs(cci,4), data = data)
> anova(ccifit)

> albuminfit <- cph(S~rcs(albumin,4), data = data)
> anova(albuminfit)

> bmifit <- cph(S~rcs(bmi,4), data = data)
> anova(bmifit)

> gfrfit <- cph(S~rcs(gfr,4), data = data)
> anova(gfrfit)

> adlfit <- cph(S~rcs(adl,4), data = data)
> anova(adlfit)
```

```
############################################################
## Transformations for ADL figure 5, appendix 1          ##
############################################################

> describe(adl)

## Linear fit ##

> adlfit.linear <- cph(S~adl, data = data)
> plot(Predict(adlfit.linear, adl = seq(0,25, by = 1)),
anova = anova(adlfit.linear), pval = T, data =
llist(adl), xlab = "GBS-ADL")

## Dichotomised at median ##

> data$adl.dichotomised <- ifelse(adl<5,0,1)
> dd <- datadist(data)
> options(datadist = "dd")
> adlfit.dichotomous <- cph(S~adl.dichotomised, data =
data)
> plot(Predict(adlfit.dichotomous), anova =
anova(adlfit.dichotomous), pval = T)

## Categorised at quartiles ##

> data$adl.quartiles <- ifelse(adl<3, 1, ifelse(adl<6,2,
ifelse(adl<10, 3,4)))
> dd <- datadist(data)
> options(datadist = "dd")
> adlfit.quartiles <- cph(S~as.factor(adl.quartiles),
data = data)
> plot(Predict(adlfit.quartiles), anova =
anova(adlfit.quartiles), pval = T)

## Two-degree polynomial ##

> adlfit.poly <- cph(S~pol(adl,2), data = data)
> plot(Predict(adlfit.poly, adl = seq(0,25,by=1)), anova
= anova(adlfit.poly), pval = T, xlab = "GBS-ADL", data =
llist(adl))

## four-knot spline ##

> adlfit.spline <- cph(S~rcs(adl,4), data = data)
> plot(Predict(adlfit.spline, adl = seq(0,25,by=1)),
anova = anova(adlfit.spline), pval = T, xlab = "GBS-ADL",
data = llist(adl))

## Log fit ##

> adlfit.log <- cph(S~log(adl+1), data = data)
> plot(Predict(adlfit.log, adl = seq(0,25,by=1)), anova =
anova(adlfit.log), pval = T, xlab = "GBS-ADL", data =
llist(adl))
```

```
#############################################################
## 5. Fitting the multivariate models                      ##
##                                                         ##
#############################################################

## The model without ADL ##

> model1 <- cph(S~age + sex + cci + rcs (hb, 4) + albumin
+ bmi + control + gfr, x = T, y = T, surv = T, data =
data)

## The full model ##

> model2 <- cph(S~age + sex + cci + rcs(hb, 4) + albumin
+ bmi + control + gfr + log (adl + 1), x = T, y = T, surv
= T, data = data)

#############################################################
## 6. Multicolinearity                                     ##
##                                                         ##
#############################################################

> cor(data)
> vif(model1)
> vif(model2)

#############################################################
## 7. Interactions                                         ##
##                                                         ##
#############################################################

## Global tests for the model without ADL ##

> z1 <- predict(model1, type = ”terms”)

> age.ia <- z1[,”age”]
> all.others <- z1[,-1]
> anova(cph(S~age.ia*all.others))

> sex.ia <- z1[,”sex”]
> all.others <- z1[,-2]
> anova(cph(S~sex.ia*all.others))
> cph(S~sex.ia*all.others)

> cci.ia <- z1[,"cci"]
> all.others <- z1[,-3]
> anova(cph(S~cci.ia*all.others))

> hb.ia <- z1[,"hb"]
> all.others <- z1[,-4]
> anova(cph(S~hb.ia*all.others))

> albumin.ia <- z1[,"albumin"]
> all.others <- z1[,-5]
> anova(cph(S~albumin.ia*all.others))
```

```
> bmi.ia <- z1[,"bmi"]
> all.others <- z1[,-6]
> anova(cph(S~bmi.ia*all.others))
> cph(S~bmi.ia*all.others)

> control.ia <- z1[,"control"]
> all.others <- z1[,-7]
> anova(cph(S~control.ia*all.others))

> gfr.ia <- z1[,"gfr"]
> all.others <- z1[,-8]
> anova(cph(S~gfr.ia*all.others))

## The full model ##

> z2 <- predict(model2, type = ”terms”)

> age.ia2 <- z2[,"age"]
> all.others.2 <- z2[,-1]
> anova(cph(S~age.ia2*all.others.2))

> sex.ia2 <- z2[,"sex"]
> all.others.2 <- z2[,-2]
> anova(cph(S~sex.ia2*all.others.2))

> cci.ia2 <- z2[,"cci"]
> all.others.2 <- z2[,-3]
> anova(cph(S~cci.ia2*all.others.2))

> hb.ia2 <- z2[,"hb"]
> all.others.2 <- z2[,-4]
> anova(cph(S~hb.ia2*all.others.2))

> albumin.ia2 <- z2[,"albumin"]
> all.others.2 <- z2[,-5]
> anova(cph(S~albumin.ia2*all.others.2))

> bmi.ia2 <- z2[,"bmi"]
> all.others.2 <- z2[,-6]
> anova(cph(S~bmi.ia2*all.others.2))

> control.ia2 <- z2[,"control"]
> all.others.2 <- z2[,-7]
> anova(cph(S~control.ia2*all.others.2))

> gfr.ia2 <- z2[,"gfr"]
> all.others.2 <- z2[,-8]
> anova(cph(S~gfr.ia2*all.others.2))
> cph(S~gfr.ia2*all.others.2)

> adl.ia <- z2[,"adl"]
> all.others <- z2[,-9]
> anova(cph(S~adl.ia*all.others))
> cph(S~adl.ia*all.others)
```

```
## Updating the models with the interactions ##

> model1 <- cph(S~age + sex * bmi + cci + rcs (hb, 4) +
albumin + control + gfr, data = data, x = T, y = T, surv
= T)

> model2 <- cph(S~age + sex * bmi + cci + rcs (hb, 4) +
albumin + control + gfr * log (adl + 1), data = data, x =
T, y = T, surv = T)

#########################################################
## 8. Proportional hazards assumption                   ##
##                                                      ##
#########################################################

> z3 <- predict(model1, type = "terms")
> model1.short <- cph(S~z3, x = T, y = T)
> ph1 <- cox.zph(model1.short, transform = "identity")
> ph1

> z4 <- predict(model2, type = "terms")
> model2.short <- cph(S~z4, x = T, y = T)
> ph2 <- cox.zph(model2.short, transform = "identity")
> ph2
> plot(ph2, var = "gfr") ##figure 6 (appendix 1) ##

#########################################################
## 9. Influential observations                          ##
##                                                      ##
#########################################################

> inf1 <- which.influence(model1)
> show.influence(inf1, dframe = data)

> inf2 <- which.influence(model2)
> show.influence(inf2, dframe = data)
> inf2

## Sensitivity analysis without influential for ADL ##

> subset <- data[-c(3,25,38,56,67,69,95,108,161),]
> attach(subset)
> S.sens <- Surv (time, status)
> sensitivity.model <- cph(S.sens~age + sex * bmi + cci+
rcs (hb, 4) + albumin + control + gfr * log(adl + 1), x=
T, y = T, surv = T, data = subset)
> sensitivity.model
> anova(sensitivity.model)
> detach(subset)
```

```
#############################################################
## 10. Relative contribution of ADL, figure 1(article) ##
##     figure 7(appendix)                                  ##
##                                                         ##
#############################################################

> plot(anova(model1), margin = "P", rm.ia = TRUE)
> plot(Predict(model1), anova = anova(model1), pval = T)

> plot(anova(model2), margin = "P", rm.ia = TRUE)
> plot(Predict(model2), anova = anova(model1), pval = T)

#############################################################
## 11. Added value of ADL, table 3 (article)             ##
##                                                         ##
#############################################################

## Likelihood ratio χ² test ##

> lrtest(model1, model2)

## Discrimination ##

> library (survC1)
> mydata <- as.matrix(data[,c("time", "status")])
> Inf.Cval.Delta(mydata, model1$x, model2$x, tau = 1428)

## NRI>0 and IDI ##

> library(survIDINRI)
> i <- IDI.INF(mydata, model1$x, model2$x, t0 = 1428)
> IDI.INF.OUT(i)

#############################################################
## 12. Internal validation                                ##
##                                                         ##
#############################################################

> validate(model1, B = 1000)
> validate(model2, B = 1000)

#############################################################
##                                                         ##
## 13. Updating the model                                 ##
##                                                         ##
#############################################################

>library(glmpath)
> mydata <- list(x = predict(model2, type = "ccterms"),
time = data$time, status = data$status)
> path <- coxpath(data = mydata)

##creating figure 8 appendix ##
```

```
> plot(path)
> plot(path, type = "aic")

## Determining the shrinkage factors ##

> lasso.factors <- path$b.predictor[path$aic ==
min(path$aic),]

## Shrinking the lasso.coefs ##

> lasso.coefs <- model2$coef

> lasso.coefs["age"] <- lasso.coefs["age"] *
lasso.factors[1]
> lasso.coefs["sex"] <- lasso.coefs["sex"] *
lasso.factors[2]
> lasso.coefs["bmi"] <- lasso.coefs["bmi"] *
lasso.factors[2]
> lasso.coefs["cci"] <- lasso.coefs["cci"] *
lasso.factors[3]
> lasso.coefs["hb"] <- lasso.coefs["hb"] *
lasso.factors[4]
> lasso.coefs["hb'"] <- lasso.coefs["hb'"] *
lasso.factors[4]
> lasso.coefs["hb''"] <- lasso.coefs["hb''"] *
lasso.factors[4]
> lasso.coefs["albumin"] <- lasso.coefs["albumin"] *
lasso.factors[5]
> lasso.coefs["control"] <- lasso.coefs["control"] * 0
> lasso.coefs["gfr"] <- lasso.coefs["gfr"] *
lasso.factors[7]
> lasso.coefs["adl"] <- lasso.coefs["adl"] *
lasso.factors[7]
> lasso.coefs["sex * bmi"] <- lasso.coefs["sex * bmi"] *
lasso.factors[2] * lasso.factors[2]
> lasso.coefs["gfr * adl"] <- lasso.coefs["gfr * adl"] *
lasso.factors[7] * lasso.factors[7]

## Updating the model ""

> lassomodel <- model2
> lassomodel$coefficients <- lasso.coefs

## Plotting nomogram, figure 2 (article) ##

> plot(nomogram(lassomodel, age = c(60,80,100), albumin =
c(15,20,30,40,45), bmi = c(15,20,25,30,35), hb =
c(50,70,90,110,150,175), interact = list(gfr =
c(27,36,51), adl, bmi, sex), lp = T, lp.at = c(-4,-
2,0,2), nint = 5, maxscale = 50))

## creating four risk groups ##
```

```
> risk.group <- cut2(as.numeric
(lassomodel$linear.predictor), g = 4)
> levels(risk.group) <- as.character(1:4)

## Kaplan-Meier plot, figure 3 (article)

> survplot(npsurv(S~risk.group, data = data), xlim =
c(0,1318), label.curves = FALSE, conf = "none", n.risk =
T, xlab = "follow-up (days)", cex.nrisk = 0.8, ylab =
"Fraction survivors", time.inc = 364, sep.n.risk = 0.03,
y.n.risk = 0, col = c(1,2,3,4), lty = 1)


### END ###
```

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract *"Cohort study" in title and abstract, page 1 and page 2*
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found *results section in abstract includes the most important findings, related to the objectives, page 2.*
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported *Rationale of improved mortality prediction [introduction, 1st paragraph] The need for improvement in studies regarding ADL and mortality [Introduction, 2nd paragraph*], page 4
Objectives	3	State specific objectives, including any prespecified hypotheses *"we aim to determine the relative importance and added value of this ADL measurement compared to clinical data, with regard to mortality prediction" [introduction 3rd paragraph], page 4*
Methods		
Study design	4	Present key elements of study design early in the paper *The methods section starts with a general description [Methods 1st paragraph], page 4. The1st paragraph of the statistical method section also describes the study design, page 6*
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection *setting, location, dates, follow-up and data collection are described on pages 4 to 6. Exposure was not dichotomous but continuous [Methods]*
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *eligibility, and methods for selection of participants [Methods, page 5] Follow-up [statistical method, page 7]*
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Not matched, Not applicable*
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable *all variables are presented in methods section [Methods, page 5 and 6] Variable transformations are presented in statistical methods section, page 7*
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group *sources of data and details of assessment are described. [Methods, page 5 and 6]*
Bias	9	Describe any efforts to address potential sources of bias *selection bias was addressed by including the group allocation variable in all analyses, as mentioned in discussion, page 13.*
Study size	10	Explain how the study size was arrived at

		"As a secondary analysis, no specific power calculation was done " [statistical methods, page 6] Before the original study a power calculation was performed, suggesting a needed sample size of 202 patients, albeit with a completely different study question.
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why *described in general in paper [statistical methods, page 6 to 9] and in detail in appendix 1*
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding *described in general in paper [statistical methods, page 6 to 9] and in detail in appendix 1*
		(*b*) Describe any methods used to examine subgroups and interactions *described in general in paper [statistical methods, page 7] and in detail in appendix 1*
		(*c*) Explain how missing data were addressed *described in general in paper [statistical methods, page 7] and in detail in appendix 1*
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *described in general in paper [results, page 9] and in detail in appendix 1*
		(*e*) Describe any sensitivity analyses *described in general in paper [statistical methods, page 7] and in detail in appendix 1*

Continued on next page

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed *eligibility is described in methods section and further details are found in the referenced original paper [methods page 5, reference 26]*
		(b) Give reasons for non-participation at each stage *This is described in the methods section and, in greater detail, in the referenced paper [methods page 5, reference 26]*
		(c) Consider use of a flow diagram *a reference to the flowchart in reference 26 is found in the methods section [methods page 5, reference 26]*
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders *descriptive data is found in table 1, page 9*
		(b) Indicate number of participants with missing data for each variable of interest *This is described in table 1, page 9 in the text of the results section, page 10 and in appendix 1.*
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount) *Follow-up time is summarised in results section, page 9 and appendix 1*
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time *number of events is reported in results section, page 9.*
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included *unadjusted estimates are presented in table 2 page 10. Multivariate estimates in figure 1, figure 2 and appendix 1.*
		(*b*) Report category boundaries when continuous variables were categorized *no categorisation or dichotomisation was done.*
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period *risks were displayed using a nomogram, figure 2*
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses *interactions are reported in statistical methods page 7, results page 11, appendix 1. One sensitivity analysis was done, with a summary in results page 11, larger detail in appendix 1.*
Discussion		
Key results	18	Summarise key results with reference to study objectives *[discussion 1st paragraph, page 12]*
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias *selection bias and Overfitting adressed in discussion page 12 and 13 and appendix 1 [methods].*
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence

		objectives, limitations, number of analysis and similar studies are discussed in discussion section, page 13.
Generalisability	21	Discuss the generalisability (external validity) of the study results *Validity (external and internal) is discussed in discussion section, page 13 and appendix 1.*
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based *Sources of funding, and roles of these, are presented in "funders", page 15.*

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.