Predictors and Prognostic Models for Early Discharge Planning of Hospitalized Acute Geriatric Patients, A Retrospective Study

Annette Hoppe

Krankenhaus Korbach

Bernhard Hoppe (✉ bernhard.hoppe@h-da.de)

Darmstadt University of Applied Sciences

Research Article

Keywords: Acute geriatric care, geriatric assessments, comprehensive treatment, predictors analysis, Logistic models, Barthel Score, Dichotomous TUG, discharge planning

Posted Date: February 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-154593/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Predictors and prognostic models for early discharge planning of hospitalized acute geriatric patients, a retrospective study

Dr. Dr. Annette Hoppe* und Prof. Dr. Bernhard Hoppe**

*) Dr. phil. nat. Dr. med. Annette Hoppe, Senior Medical Consultant, Stadtkrankenhaus Korbach, Enser Str. 19, D-34497 Korbach, A.Hoppe@krankenhaus-korbach.de

**) Prof. Dr. phil. nat. Bernhard Hoppe, University of Applied Sciences, Birkenweg 10, 64295 Darmstadt, Bernhard.hoppe@h-da.de

Corresponding Author: Bernhard Hoppe

Keywords: Acute geriatric care, geriatric assessments, comprehensive treatment, predictors analysis, Logistic models, Barthel Score, Dichotomous TUG, discharge planning

Abbreviations:

☐ ADL	Activities of daily living
☐ AUC	Area under the Curve
☐ BS	Barthel-Score
☐ BA	Barthel-Score at admission
☐ BD	Barthel Score at discharge
☐ CDT	Shulman's Clock-Drawing-Test
☐ GDS-15	Geriatric depression scale
☐ HG	Handgrip averaged over both hands in kPa
☐ M	Mean
☐ MMSE	Mini-Mental State Examination
☐ ROC	Receiver Operator Characteristics
☐ SD	Standard deviation
☐ TUG	Timed-up and Go-Test
☐ TUGd	Timed-up and Go-Test at discharge
☐ WD	Walking distance in m

Abstract

Motivation: When treated for an acute disorder, older adults are vulnerable for functional losses and the need of care after discharge. In a specialised geriatric ward, patients get a comprehensive treatment complementary to medical care in order to maintain and improve mobility and activities of daily living (ADL) to facilitate the return to domesticity. The aim of this paper is to identify the relevant predictors for the impact of geriatric treatment on the status at discharge, which are then used in logistic models to predict a patient's potential to reach a certain level of independence during treatment.

Method: In a retrospective cohort study with 580 patients, we analysed the impact of acute geriatric early rehabilitation on the functional outcome after treatment. As a sufficient improvement of ADLs and mobility we defined as a suitable endpoint at least 60 Barthel Points (ADL) and the ability for „Timed-Up-and-Go-Test“(TUG) when discharged from acute hospital care. To identify relevant predictors in the set of the screening assessments at admission we used linear and logistic regressions as well as odds-ratios. Multivariate logistic models are used to predict the probability that a patient reaches the endpoint. Their predictive quality is tested on additional 120 test patients from a different cohort.

Results: Statistical analysis shows that all patients improved during early rehabilitation significantly in ADLs and the physical function (TUG). Barthel-Score, walking distance and handgrip on admission are the strongest predictors for the outcome after geriatric treatment. Logistic models predict the outcome correctly in 70% to 80% of the cases. These models once established for a certain cohort of patients can be applied with similar accuracy to different sets of patients as well. Clinical condition, the medical treatment before admission, length of hospitalization, age or gender have no predictive quality.

Discussion: We were able to show that all patients benefit significantly from early rehabilitation in an acute geriatric ward. Only a few assessments on admission related to physical function are sufficient to indicate the functional outcome after geriatric treatment. Logistic models based on these predictors are reliable with a generic predictive quality for the expected level of independence at discharge. This facilitates early discharge planning.

Trial registration: The study was retrospectively registered on 2nd July 2018 by the ethic commission of the hospital und filed under registration number (MG1/569/770/2019).

Background

Older adults are the largest group of hospitalized patients. In western countries up to 50% [1] of the admitted persons are 65 years or older. Most older adults live at home (74% in Germany and 67,5% in OECD25 [2]). However, older patients in hospitals suffer severe risks of functional decline. Early rehabilitation of the elderly along with the clinical treatment in specialized acute geriatric wards, labeled often as Acute Care for the Elderly units (ACE), minimizes these risks. Such a treatment is an important and useful strategy, to facilitate a return to their previous social and domestic live [1,3,4]. However, early rehabilitation patients typically need support or even institutional care after discharge. Hence, discharge planning is an important task, which should start as early as possible. To aid this process, a prognostic tool, which quantifies the potential of a patient already based on geriatric assessments at admission would be helpful.

To explore the potential of a patient we need a useful and efficient outcome measure. There is a wide variety of outcome measures reported in the literature [5], where mobility and the Barthel score for ADLs were found to be most sensitive for the benefits of early rehabilitation in acute care. Hence, we use as a first outcome target a Barthel-Score of minimum 60 points, which is the approved level of independence for acceptance in a further rehabilitation unit. As independence needs a certain capability to stand up and walk, the second outcome target is the ability to perform the Timed-Up-and-Go-Test (TUG). Both outcome targets form a meaningful endpoint. To identify strong predictors for a positive outcome we analyze the scores from standard geriatric assessments at admission. With predictors showing the largest effect-sizes logistic models are formed, to predict the probability to meet the outcome target successfully. These prognostic models are tested on different cohorts of patients to analyze their predictive power.

Organization of the paper is as follows. At first, we describe the procedures and participants of this retrospective study, the geriatric assessments and give an overview of the statistical methods and a descriptive statistical listing of the data. As our endpoint contains Barthel-Score and TUG as outcome measures, we first analyze dependencies of BD and TUGd on admission assessments separately. We extend this analysis to the combined output measure as a suitable endpoint and develop a predictive model for prognosis of the patient's potential reaching the endpoint. We describe the overall model quality using sensitivity, specificity and the area under the ROC curve. To test the quality of the models, we apply the models to data of a different cohort of test-patients.

Methods

Procedures and Participants

Participants in this retrospective cross sectional study were 700 randomly chosen patients of the acute geriatric ward of a medium sized hospital in a rural area in Germany from 2009-2019. Data of 580 patients are analyzed and 351 of them with a BA less than 60 are used for model development. Another group of 120 patients, treated in 2018 and 2019, form the test cohort.

Patients are typically aged > 70 years, have multiple chronic illnesses and are admitted because of an acute disorder. They were transferred to the geriatric ward either from the surgical or internal ward of the hospital or other hospitals nearby without a geriatric unit. The disorders treated were internal (apoplexy, cardio-vascular disorders, diabetes, pneumonia, infections) or surgical (fractures of femoral

neck, pelvis, vertebral bodies and humerus, treated surgically or conservatively) or convalescents after abdominal or cardiac surgery.

During their stay (typically 14 days) patients get, complementary to acute medical care, a comprehensive training in order to maintain and improve mobility and activities of daily living. The rehabilitative training comprises intense activating daily care, physiotherapy, occupational therapy, physical exercise, cognitive training and if needed speech therapy and psychological consulting. The training is tailored to meet the specific individual deficits of a patient. Ten multidisciplinary therapeutic training sessions are scheduled per week. Staff-members (nursing, therapists and doctors) in the geriatric ward have a special geriatric qualification and meet regularly to discuss the development of the patients.

Randomization

Participants in this study were randomly chosen, selected by picking arbitrary hospital registration numbers of patients admitted to the geriatric ward from 2009-2019. Data of 580 patients form the basic cohort used for statistical analysis and 351 datasets with BS < 60 were used for model development. Other 120, randomly chosen patients with BS < 60 form a test cohort, to validate the predictive power of the model. To compensate for bias and outliers in the randomly chosen cohorts, bootstrap resampling was applied to derive 95%-confidential intervals.

Measures

For the whole set of treatment, the hospital gets a fixed remuneration under certain conditions: Specially trained staff, adequate equipment and furnishing of the ward, size of the hospital rooms and specific procedures. In particular, the testing of mobility, ADLs, cognition and emotion on admission and ADLs and mobility at discharge is mandatory. The specific tests were selected by the hospital and then accredited by the federal geriatric association.

The following tests were applied at admission: Barthel-Score (**BS**), Timed-up and Go-Test (**TUG**), walking distance (**WD**) and handgrip strength (**HG**), Shulman's Clock-Drawing-Test (**CDT**) and Mini-Mental State Examination (**MMSE**) and the geriatric depression scale (**GDS-15**). Tests for physical function (TUG, HG, WD) and BS were repeated before discharge.

The Barthel-Score (BS) is a marker for the performance in ADL. A higher score is associated with a greater likelihood of being able to live with a certain degree of independence. According to ICD-10, Barthel-Scores are clustered in 5 intervals. A score 60-75 indicates a medium sized impairment. In the three intervals below 60 there is growing dysfunctionality with falling Barthel-Scores. There is none or only a slight impairment for scores from 80 to 95.

TUG-testing results indicate fall risks and measure the impairment of mobility by taking the time in seconds required to stand up from a chair, followed by walking 3 meters, turn around walk back and sit down again. The TUG-performance is measured in seconds and hence a quantitative indicator for physical functioning as frailty and falling risk but is also a useful indicator for cognition [11, 12, 22, 25, 29]. A TUG time more than 12 seconds is predictive for future falls [24].

Hand grip strength correlates positively with overall physical performance and has a predictive validity for decline in cognition and mobility [6, 13, 14]. Handgrip was measured in kPa using the Vigorimeter (KLS Martin), which is as reliable as the JAMAR Dynamometer [15]. In the present study, we consider the averaged pressure of both hands.

The walking distance (WD) is an indicator for physical function, cognition and mental status as well as their interrelationship [16, 17]. This test accounts for three important aspects of overall functioning. We measure the untimed walking distance (WD) in meter.

For the assessment of cognition, we use the Mini-Mental State Examination (MMSE) [18] and Shulman's Clock-Drawing-Test (CDT) [19, 20]. MMSE is a screening test for dementia with a maximum score of 30 points. The threshold for no or negligible impairment are 24 points. There are three categories for mild (19–23 points), moderate (10–18 points) and severe deficits for less or equal 9 marks. For performing in the Clock-Drawing-Test patients have to draw the dial of an analogous watch with the watch hands set to certain time: 10 past 11. The outcome is assessed with a 6 point inverted scale: 3 or more points indicate cognitive impairment.

The score for the Geriatric Depression Scale (GDS) is determined via individual self-assessment of the patient [21]. 15 questions have to be answered by yes/no. From a scoring grid with one point per answer, the severity of depression can be assessed. Below 5 points there is no depression, a mild depression is to be expected between 5 to 10 points. 11 to 15 points indicate a severe depression. This test is not possible in the case of dementia.

Besides the screening tests the following characteristics - taken from the hospital records – were included for each patient in the analysis: age, gender, duration of stay in days and the referral from surgical or medical department.

Endpoint and Outcome measures

Heldmann and coworkers [5] analysed outcome measures in a meta-analysis for acute geriatric early rehabilitation. ADL-capacity (BS) is a typical indicator, substantiated in many cases by mobility criteria.

In our view independent or moderately assisted living after treatment, requires a certain capability in ADLs plus a basic mobility e.g. transfer from a passive state (i.e. sitting) to active movements (i.e. walking) and vice versa, while maintaining a low falling risk. As a sufficient improvement of ADLs and mobility, we define as a suitable endpoint at least 60 Barthel-Points plus the ability for an untimed TUG.

A drawback of the timed TUG-test for hospitalized patients is that many are not able to stand up and therefore are incapable to perform the timed test. Hence a certain improvement in seconds at discharge is an outcome measure with a large floor effect. Dichotomous tests (TUG possible or not) also reported in literature [12], avoid this effect. It was shown, that the timed as well as the dichotomous test are significantly associated with functional performance. Patients found able to do the TUG had a lower fall rate as those who failed. We therefore use the dichotomous test as a suitable outcome measure and monitor the TUG-time as well.

Statistical methods

We use SPSS 27 to perform the statistical analysis [23] of the assessment data. A significance-level $p < 0.05$ was used throughout. The magnitude of a certain effect is measured using standardized effect-size parameters r, f, V : $r > 0.5, f > 0.4$ and $V > 0.5$ indicate a strong effect.

The assessment data may exhibit deviations from a normal distribution. For parametric test e.g. t-tests non-parametric tests (e.g. Wilcoxon or Mann Whitney U tests) validate statistical findings. The chi-squared test is used to examine statistical properties of categorical variables and odds-ratios to quantify the impact of dichotomous measures (e.g. gender) on the outcome of the geriatric treatment. With Bootstrapping using 1000 samples, we obtain 95%-ranges for the statistical findings.

Linear and logistic regressions were performed to examine the impact of the screening parameters (predictors) taken at admission (BA, HG, WD, MMSE etc.) on the outcome at discharge. Linear regression analysis results not only in optimal fits of the data with a straight line but comprises correlation of predictors with endpoints as well. Logistic regression defines a model for the probability P ($0 \leq P \leq 1$) to reach a certain condition, here to meet the endpoint. Using $P = 0.5$ as the classification

cutoff, a binary classification is possible i.e. a go/no-go-decision. P is a logistic function of the metric assessment scores weighted with coefficients fitted to the data. For a bivariate analysis e.g. $P(BA)$ BA is the only variable considered.

Logistic regression models for a probability once established for a given set of data, may be applied to other data sets. If the probabilities predict the outcome in such test sets with comparable accuracy, then the models obtained from the initial data have an inherent generic predictive quality. To investigate this, we compared the clinical outcomes to the prognosis of the models for the initial data set and for 120 different test-patients. Quality measures are sensitivity and specificity of the P -models using $P = 0.5$ as the classification cutoff and the ROC-Curves (Receiver Operating Characteristics). The ROC-curve evaluates a model by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity) of the model-predictions with varied decision threshold settings (i.e. P -values). A reference for no predictive quality is a random classifier giving a 45°-line in the ROC-diagram. Specifically, the area under ROC-curves (AUC) is a quantitative quality criterion. The range is between 0 and 1 and the larger the area, the better [30]. A suitable index-number for the predictive quality is AUC minus the area under the 45°-line multiplied by 2. This defines the *Gini-Index* which varies between 0 and 1. The best discrimination threshold yields a maximum number of true positives and the minimum number of false positives. This threshold maximizes the *Jouden-Index* which is sum of sensitivity plus specificity minus 1. In the ROC curve the best discriminator corresponds to the point with the largest distance from the 45° line and maybe determined by placing the 45°-tangent to the curve.

Patient data

The basic cohort with 580 patients consists of 426 women, 154 men. 291 of them had internal diseases 289 surgical problems (84 of them were treated conservatively). Screening assessments at admission, are listed Table 1.

Scores and statistics	Valid measures	Range	$M \pm SD$	Median
Barthel score (BA)	580	0 .. 95	49.8 ± 17.4	50
GDS-15	543	0 .. 14	3.6 ± 2.7	3
MMSE	551	6 .. 30	24.3 ± 5.0	26
CDT	515	1 .. 6	3.1 ± 1.3	3
TUG in seconds	194	8 .. 80	26.7 ± 9.2	27
HG re/li (averaged) in kPa	541	0 .. 82	29.3 ± 16.7	28
WD in m	468	1 .. 150	25.0 ± 25.9	30
Age in years	580	64 .. 97	82.7 ± 6.3	83
Weight in kg	575	36 .. 145	71.7 ± 15.9	70
Hospitalization in days	580	7 .. 46	18.2 ± 3.9	18

Table 1: Patient characteristics recorded at admission, range, median, mean values (M) and standard deviations (SD).

Predictors for Barthel-Score or TUG at discharge

The endpoint of this study consists of two components: BD / TUGd. As a first step, we investigate the predictors for BD and TUGd independently to quantify the impact of the geriatric early rehabilitation scheme on ADLs and mobility respectively.

Barthel Score

Patients BA score was categorized according to the five ICD 10 groups. The improvement of the score during hospitalization is determined using the difference of means ($M\Delta$) before and after treatment in each category and for all $n = 580$ patients (Fig. 1). A paired sample t -test for each group before and after treatment is performed and the results are confirmed by a non-parametric Wilcoxon test, Tab. 2. The increase in BS is significant with a strong effect size in all groups $0.78 < r < 0.91$. BD minus BA is around 20 points, i.e. on the average patients improve their ADL-score by one ICD10 group. The groups 20 to 35 and 40 to 55 have the highest benefit from early geriatric rehabilitation.

ICD10-group	$M\Delta$: $M(BD)-M(BA)$ (95% CI)	SD of $M\Delta$	p	r	p	r
			t-test		Wilcoxon-test	
BA ≤ 15 $n = 27$	15.3 (10.2- 20.5)	9.7	0.00	0.78	0.00	0.76
BA: 20 to 35 $n = 112$	21.6 (18.9 - 24.4)	12.3	0.00	0.83	0.00	0.83
BA: 40-55 $n = 212$	22.4 (20.7 - 24.1)	10.3	0.00	0.87	0.00	0.83
BA: 60-75 $n = 224$	19.1 (18.0 - 20,2)	7.5	0.00	0.91	0.00	0.85
BA: 80-95 $n = 5$	11.0 (2.9 - 19.1)	6.6	0.02	0.88	0.04	0.91
All patients $n = 580$	20.6 (19.6 - 21.5)	19.0	0.00	0.87	0.00	0.83

Table 2: Mean differences ($M\Delta = M(BD)-M(BA)$) in bold letters and standard deviation (SD) before and after treatment with 95% confidence intervals in the ICD10 categories and for the full cohort.

Fig. 1 Barthel Scores at admission (BA) and before discharge (BD) and BS–improvement with standard deviations indicated by black lines

Predictors for BD

To find the most important predictors for BS-improvement we examined all metric scores at admission for all 580 patients, (except CDT, which is categorical), with bivariate linear regressions (Table 3). F -tests prove significance for the linear model and t -tests significance for the coefficients (i.e. constant and slope of the linear fit). R^2 , the measure of determination, is the squared Pearson correlation

coefficient R and f is Cohens effect size. The table lists significant models only with a large enough F -value for $p < 0.05$.

As expected, the Barthel score at admission (BA) has the strongest impact on the score at discharge (BD). BA correlates strongly with BD ($R = 0.82$) and Cohen's effect size is $f = 1.424$. HG, WD and MMSE follow in descending order: $R = \{0.44, 0.40, 0.39\}$; $f = \{0.492, 0.433, 0.425\}$. Age, GDS-15 and TUG-time are weak predictors, whereas weight and duration of treatment are non-significant ($p > 0.2$) and hence are not included in Table 3. In particular age is weakly correlated with BD: $R = 0.16$. The linear regression models for BD are shown in Fig. 2.

Predictors for BD	$F(df,n)$	R^2	Constant	Slope	f
BA	$F(1,576)=1168$	0.67	$19.2 < \mathbf{22.12} < 25.1$	$0.9 < \mathbf{0.97} < 1.0$	1.424
HG in kPa	$F(1,569)=138$	0.195	$52.0 < \mathbf{55.33} < 58.8$	$0.4 < \mathbf{0.53} < 0.6$	0.492
WD > 0 in m	$F(1,466)=87$	0.158	$65.1 < \mathbf{67.43} < 69.6$	$0.20 < \mathbf{0.26} < 0.31$	0.433
MMSE	$F(1,548)= 99$	0.153	$25.1 < \mathbf{33.96} < 43.0$	$1.2 < \mathbf{1.55} < 1.9$	0.425
TUG in s	$F(1,192)=6,9$	0.035	$85.8 < \mathbf{90.48} < 94.9$	$-0.4 < \mathbf{-0.23} < -0.1$	0.190
GDS	$F(1,539)=11.1$	0.020	$73.3 < \mathbf{75.85} < 78.4$	$-1.6 < \mathbf{-1.02} < -0.4$	0.146
Age in years	$F(1,577)= 15$	0.026	$92.2 < \mathbf{113.43} < 135.9$	$-0.8 < \mathbf{-0.53} < 0.3$	0.161

Table 3 Linear regression parameters ($p < 0.05$ for all predictors listed) for bivariate analysis of BD. 95%-ranges for slope and constant are given for each significant regression. Bold are the nominal values of the regression coefficients.

Fig. 2 Linear regression models for BD vs. predictor ranges in %: BA (solid line), MMSE (dashed line), HG (dash dotted) and WD (dotted)

The figure shows that BA has the highest leverage on the Barthel score at discharge. The other predictors with identical but almost smaller slopes have a moderate impact on BD and they need a large vertical intercept to interpolate between the measured data. This finding is in perfect accordance with the effect sizes and the Pearson correlation coefficients, see Tab. 2.

TUG-test at admission and discharge:

In acute geriatric care a TUG test is a challenge for patients. Only 194 patients were able to do the TUG test at admission and 406 at discharge. The mean TUG time at admission was 26.7 ± 9.2 s. Only six

patients had a TUG time under the 12 seconds limit [24], the cut-off point for normal mobility. The TUG time at discharge of the 194 patients with a timed TUG at admission, was 17.9 ± 7.5 s, which is a significant improvement (paired t -test: $t = 19.4$, $p = 0.000$, $r = 0.82$; Wilcoxon test: $p = 0.000$, $r = 0.85$). The number of patients with TUG under 12s increased from 6 to 47 patients.

386 patients were unable to do the TUG at admission but 215 of those patients could do the TUG at discharge with a TUG time of 24.8 ± 9.8 s. 10 of them had a TUG time under the 12 seconds limit.

Dichotomous variable: Logistic Regression

As TUG times at admission as well as on discharge are typically well above the cut off value of 12 seconds and as only 33% of the patients were able at admission to do the test, the dichotomous version of the test (TUG possible or not) is the appropriate choice. Bivariate logistic regressions are applied to identify strong predictors for such dichotomous outcome-measures. χ^2 -tests prove the significance for the logistic model. The coefficient of determination is Nagelkerke's R^2 , from which Cohens f is obtained. Table 4 lists the results for the significant models ($p < 0.05$).

Predictor for TUGd	$\chi^2(df=1)$	R^2	β_0	β_1	f
BA	$\chi^2 = 221.9$ $n = 579$	0.451	$-4.31 < \mathbf{-3.57} < -2.91$;	$0.08 < \mathbf{0.10} < 0.11$;	0.906
WD > 0m	$\chi^2 = 79.1$ $n = 468$	0.245	$-0.48 < \mathbf{-0.02} < 0.40$;	$0.04 < \mathbf{0.06} < 0.09$;	0.570
HG in kPa	$\chi^2 = 58.2$ $n = 572$	0.138	$-0.69 < \mathbf{-0.32} < 0.07$;	$0.03 < \mathbf{0.05} < 0.06$	0.400
MMSE	$\chi^2 = 33.6$ $n = 551$	0.085	$-2.84 < \mathbf{-1.74} < -0.75$;	$0.04 < \mathbf{0.08} < 0.13$;	0.305
Age	$\chi^2 = 8.0$; $n = 577$	0.019	$1.83 < \mathbf{4.22} < 6.83$;	$-0.71 < \mathbf{-0.04} < -0.01$	0.139

Table 4 Nominal logistic regression parameters β_0 and β_1 in bold letters for bivariate analysis of impact of predictors on TUGd. 95%-ranges for β_0 and β_1 are given for each predictor. All predictors listed are significant, $p < 0.05$.

The important predictors for TUGd with a strong effect size $f \geq 0.4$ are BA ($f = 0.906$), WD ($f = 0.570$), HG ($f = 0.400$). MMSE ($f = 0.305$) and age ($f = 0.139$) establish significant models but the effect size is medium or small respectively. Note that WD and HG change positions in the hierarchical order as compared to the linear BD-regression. In particular, the small effect size of age indicates that age is not very important for performing the TUG at discharge. Weight, GDS and hospitalization in days have a non-significant effect and hence are not considered.

Predictors for BD ≤ 60 and TUGd (endpoint) for patients with BA < 60

The endpoint of this study is the joint performance in metric BS (BD ≤ 60) and in the dichotomous TUG (TUGd possible or not). Reaching the endpoint successfully is meaningful only for patients whose BA is less than 60 to avoid biasing. This applies for 351 of the total 580 patients.

TUG and BS in the BA < 60 group

All metric scores of these 351 patients at admission are listed in Table 5. Comparing Tab. 5 and Tab. 1 shows that major differences are (aside BA) found in WD, HG and particular in TUG: The ability to perform the TUG in the full cohort is $194/580 = 33\%$ and in the BA < 60 group $50/351 = 14\%$ only. In total 164 of the 351 patients are reaching the endpoint but 184 (52%) could do the TUG at discharge.

It should be noted that 20 patients who could perform the TUG at discharge, failed to reach a BD ≥ 60 and 58 out of 222 patients with BD ≥ 60 failed to perform TUGd.

Scores and statistics	Valid measures	Range	$M \pm SD$	Median
BA < 60	351	0 .. 55	38.9 ± 13.4	40
GDS-15	317	0 .. 14	3.8 ± 2.7	3
MMSE	326	5 .. 30	23.3 ± 5.4	25
CDT	297	1 .. 6	3.4 ± 1.3	3
TUG in seconds	50	16 .. 80	29.5 ± 10.1	29
HG in kPa	318	0 .. 66	25.0 ± 15.6	24
WD in m	248	1 .. 100	22.7 ± 21.8	15
Age in years	351	65 .. 97	83.4 ± 6.4	83
Weight in kg	347	36 .. 145	71.4 ± 16.1	69
Weight female in kg	250	36 .. 145	68.2 ± 15.4	66
Weight male in kg	97	49 .. 112	79.9 ± 14.7	80
Hospitalization in days	351	7 .. 46	18 ± 4	18

Table 5: Assessments for patients with BA below 60 points

The 50 patients who were able to do the TUG test at admission have a mean TUG time of 29.5 ± 10 s, but none had a TUG < 12 s. After treatment 48 of the 50 patients still could do the TUG and their TUG time improved to 20.7 ± 9.2 s but was well above the cut off value of 12 s. From 301 patients who were not able to perform the TUG test initially, 136 could do the test at discharge with a TUG time of 26.3 ± 9.7 sec. This difference of TUGd between patients who could perform the test at admission to those who were unable is significant: (26.3s / 20.7s) (t-test for unpaired groups $p < 0.01$, $r = 0.36$; Mann Whitney-U-test, $p = 0.000$, $r = 0.28$). In total only 8 patients reached a TUG below 12s. 167 patients failed to perform the TUG at discharge.

In this group (BA < 60) patients improve their BS-scores by 21.6 ± 13.4 (n=351) and 129 still had a BD less than 60. Patients with a successful TUG at discharge improve their BD significantly higher than those who were unable to do the TUG at discharge (BD = 27.2 ± 11.4 and BD = 15.4 ± 12.7). This difference is significant (unpaired t-test $p = 0.000$, $r = 0.45$ or Mann-Whitney-U-Test $p = 0.000$, $r = 0.47$).

Patients meeting both endpoint criteria enhance their BS-scores significantly higher by 28.7 ± 9.9 (n=164) than those, who fail to do so: 15.3 ± 13.0 . A t-test for unpaired groups results in $p = 0.000$, $r = 0.50$ and the Mann Whitney-U-test gives $p = 0.000$, $r = 0.69$. TUG-time for patients reaching the endpoint is $24.3s \pm 9.6s$. 20 patients with TUGd but BD < 60 had a TUG = $29.4 \pm 10.9s$. Though the difference of means is large, the t-Test is non-significant: $p = 0.058$.

Bivariate Logistic Regression

Table 6 lists the results of logistic regressions for meeting the endpoint based on screening data taken at admission. The important predictors with a strong effect size $f \geq 0.4$ are BA ($f = 0.653$) and WD ($f = 0.443$). HG ($f = 0.313$) and MMSE ($f = 0.305$) only show a medium effect strength. Compared to the analysis of the previous chapter, the important predictors have the same hierarchical order as for TUGd alone. Age, weight, GDS and hospitalization in days have a non-significant effect ($p > 0.1$) and hence are not reported in the Table.

Predictor for endpoint	$\chi^2(df=1)$	R^2	β_0	β_1	f
BA < 60	$\chi^2 = 85.67$ n = 351	0.299	-4.88 < -3.81 < -3.00;	0.07 < 0.09 < 0.12;	0.653
WD > 0m	$\chi^2 = 31.7$ n = 248	0.164	-0.90 < -0.53 < 0.13;	0.043 < 0.04 < 0.06;	0.443
HG	$\chi^2 = 23.7$ n = 343	0.089	-1.4 < -0.97 < 0.58;	0.02 < 0.03 < 0.05	0.313
MMSE > 0	$\chi^2 = 20.4$ n = 326	0.081	-3.61 < -2.40 < -1.43;	0.06 < 0.10 < 0.15;	0.305

Table 6 Logistic regression parameters ($p < 0.05$ for all predictors listed).

Odds-Ratios for nominal predictors

Some of the predictor variables are binary: gender, main diagnosis, admission from the internal or surgical ward with conservative treatment or surgery. To quantify their influence, we investigated Odds-Ratios (Tab. 7, Fig. 3). Moreover, for rounding up the above findings, metric predictors are transformed into binary ones by using appropriate cut-offs taken from descriptive statistics of the 351 patients (Tab. 5). For CDT, GDS and MMSE the threshold is set to the cut-off between impairment and no impairment

Predictor A/B	$\chi^2(df=1);$ n_A/n_B	Chance A	Chance B	Odds = Chance A/ Chance B	p	V
BA χ 40 / < 40	59.651 212/139	1.753	0.271	3.973 < 6.469 < 10.630	0.000	0.414
WD (m) > 20/ $\leq$ 20	23.391 120/128	2.750	0.750	2.117 < 3.651 < 6.582	0.000	0.307
HG (kPa) χ χ 25/ < 25	12.893 164/151	1.485	0.687	1.396 < 2.191 < 3.348	0.001	0.193
MMSE χ 24/ < 24	14.626 197/149	1.318	0.542	1.535 < 2.430 < 3.848	0.000	0.212
Internal/surgical	4.470 148/203	1.162	0.732	1.033 < 1.585 < 2.433	0.035	0.113
Cons./surgery	4.275 60/143	1.143	0.602	1.030 < 1.898 < 3.496	0.039	0.146
CDT < 3/ χ 3	3.723 77/219	1.406	0.840	0.989 < 1.673 < 2.830	0.054	
Female/male	2.212; 252/99	0.969	0.678	0.89 < 1.43 < 2.29	0.137	
Weight (kg) > 71 / $\leq$ 71	5.244 152/195	1.202	0.730	1.073 < 1.648 < 2.531	0.022	0.123
Weight men (kg) > 79/ $\leq$ 79	2.644 52/45	0.529	1.048	0.221 < 0.505 < 1.155	0.104	
Weight women > 68 / $\leq$ 68	1.780 108/142	1.204	0.855	0.851 < 1.408 < 2.328	0.182	
Age < 85/ χ 85	0.340 198/153	0.942	0.829	0.742 < 1.175 < 1.736	0.560	
GDS < 5/ χ 5	0.017 219/98	0.991	0.960	0.641 < 1.032 < 1.661	0.896	

Table 7 Odds-ratios (95%-ranges) for the binary outcome variable TUGd feasible and BD $\leq$ 60. Cramer's V is listed for all significant findings. $V \leq 0.5$ indicates a strong effect. Medium sized effects are in the range $0.50 \leq V \leq 0.3$. Smaller Vs correspond to a weak effect.

Larger effect sizes are found for binary versions of BA, WD, MMSE and HG. CDT is on the verge of being significant, and age above or below 85 plays a non-significant role for the outcome of the treatment. Only the method of treatment in the surgical ward has a minor impact on the outcome. Gender is also not significant.

Fig. 3 Odds-ratios of predictors for reaching the endpoint calculated for 351 patients with BA < 60. Lines indicate the 95% interval, filled dots are ratios with a significant difference in chances for both groups.

Age as a confounder?

To rule out age as a confounder for the dominant predictors, we analyze bivariate linear regressions for BA, WD, MMSE and HG with age as the input variable. Age is a significant predictor only for handgrip but with a small effect size. See Table 8.

Age as predictor	F, p	R ²	A; p	b, p	f
HG	F(1,341)= 4.931; p = 0.027	0.0142	49.199 p=0.000	-0.290; p=0.027	0.120
BA	F(1,349)= 0.394 p = 0.531	0.0012	44.707 p=0.000	-0.070 p=0.531	0.034
MMSE	F(1,326)= 2.133 p = 0.145	0.007	29.033 p=0.000	-0.068 p=0.145	0.084
WD	F(1,247) = 0.509 p = 0.476	0.002	35.343 p = 0.048	-0.152; p= 0.476	0.045

Table 8 Linear regression for HG, MMSE, BA, WD with age as the predictor. Non-significant p values in bold letters.

Summary of bivariate analysis

The above investigations showed that only a few of the assessments at admission exhibit a significant impact with a relevant effect size on the improvement of BS, TUGd or the endpoint. Fig. 4 summarizes the effect sizes found in bivariate regression analysis. Important predictors are BA, WD, HG and MMSE.

Fig. 4 Effect sizes of significant predictors in bivariate regressions for BD, TUGd (580 patients) and reaching the endpoint (351 patients with BA < 60 at admission) as outcome variables.

Note that age, weight, GDS, gender, length of hospitalization are not significant for the endpoint. We use the important metric predictors to develop prognostic models in the following chapter.

Prognostic Models

Logistic regression models with appropriate predictors as variables yield probabilities (P) for reaching the endpoint. In this way, patients with the highest rehabilitation potential maybe identified. Logistic models with three, two and a single one of the strong predictors identified above were derived. Table 9 shows the model-parameters for the multivariate models. Parameters for bivariate models may be taken from Table 6. Fig. 5 shows as a sample a 2D-plot of the probability function $P(\text{BA}, \text{WD})$. We see that BA with the largest effect size creates a larger curvature. This is in accordance with the decreasing effect size.

Fig 5 2D-plot for $P(BA, WD)$

Predictor for endpoint	χ^2	R^2	$\beta_o; [exp(B)]$	BA; [exp(B)]	HG; [exp(B)]	WD; [exp(B)]	f
BA, WD, HG n = 247	$\chi^2(d=3) = 63.48;$	0.305;	-5.3<- 3.77 <-2.5 [0.023]	0.04< 0.07 <0.10 [1.070]	0.00< 0.021 <0.04 [1.021]	0.02< 0.03 <0.06 [1.021]	0.662
WD, HG n = 247	$\chi^2(d=2) = 40.06;$	0.201;	-1.9<- 1.178 <-0.6 [0.308]		0.01< 0.027 <0.05 [1.027]	0.02< 0.038 < 0.06 [1.039]	0.502
BA, HG n = 343	$\chi^2(d=2)= 89.02;$	0.305;	-5.0<- 3.983 <-3,1 [0.019]	0.06 < 0.085 < 0.11 [1.089]	0.002< 0.019 <0.04 [1.019]		0.662
BA, WD n = 248	$\chi^2(d=2)= 60.50;$	0.291;	-4,84<- 3.44 -2,44 [0,032]	0,05< 0.072 <0,10; [1.075]		0.02< 0.033 <0.06 [1.034]	0.641

Table 9 Significant multivariate logistic regression for reaching successfully the endpoint. Adding MMSE as a forth predictor (BA, WD, HG, MMSE) results in a non-significant model (not listed).

Another more practical way of visualizing probabilities are contour-plots (Fig. 6). For example, the probability P to meet the endpoint with a certain HG and WD at admission maybe taken from the intersection of the corresponding vertical and horizontal lines. If for instance HG = 10 kPa and WD = 40m the probability for a reaching the endpoint is 66%, see Fig. 6. To use the three dimensional model $P(BA, WD, HG)$ a graphic evaluation is not possible and P has to be obtained by computation. The formula reads

$$P(BA, HG, WD) = \frac{1}{1 + \exp(-(-3.77 + 0.07 \cdot BA + 0.021 \cdot HG + 0.03 \cdot WD))}$$

Fig 6 Contour plot for $P(WD, HG)$ with contour lines with a 0.1-spacing. Red dashed line is the $P = 0.5$ line, which is used for classification. The intersection of $WD = 40m$ and $HG = 10kPa$ determines the probability to reach the endpoint.

Test of Models

As a first test we determined sensitivity, specificity and correct classification rate using $P = 0.5$ as the classification threshold (Table 10). If $P > 0.5$ then we expect the patient to reach the endpoint otherwise not. Table 10 shows the results for patients with $BA < 60$, left: 351 patients used for model derivation and right: 120 additional patients from the test cohort, whose screening parameters at admission were not used for model derivation.

Predictor(s)	Sensitivity	Specificity	Correct classification	Sensitivity	Specificity	Correct classification
	351 BA < 60 Patients			120 Testpatients BA < 60		
BA	67	70	68	60	84	70
HG	51	71	61	69	62	66
WD	67	66	66	83	64	75
BA, WD	79	56	69	86	70	79
BA, HG	74	62	68	73	76	74
WD, HG	74	54	66	90	64	79
BA, WD, HG	80	59	72	90	70	82

Table 10 Sensitivity, specificity and total correct classification in % for logistic models using different predictor combinations.

Obviously, qualitatively similar classification rates are achieved for the test-set of 120 patients as for the cohort of 351 patients.

A systematic evaluation of model quality is done by using ROC-curves, AUCs and the relevant indices for the test-set and the basic cohort, Tab. 11. Fig. 7 shows these curves for the models with the best Gini-Indices ($P(BA, WD, HG)$, $P(BA, WD)$ and $P(WD)$). We see that in both cases the three predictor-model has the largest area under the curve and hence gives the best classification results of the three models. The optimal cut-off (Jouden-Index) is found by inspecting the coordinates of the ROC-curves.

The index corresponds to the 45°-tangent to the ROC-curve shown in Fig. 7 for the three predictor model. In both plots there is not a singular touching point but a certain range is possible to place the tangent. The best cutoff-point for symmetrical sensitivity and specificity (351: 72% to 72% or 120: 78% to 78% respectively) is indicated in the ROC-curves and is found in both cases from the coordinates at $P = 0.57$ or $P = 0.63$.

Predictors	351 Patients			120 Test-Patients		
	AUC	Gini	Co:Jouden	AUC	Gini	Co:Jouden
BA	0.636 < 0.704 < 0.773	0.409	0.46:1.40	0.729 < 0.809 < 0.889	0.618	0.46:1.45
HG	0.560 < 0.632 < 0.705	0.265	0.48:1.21	0.651 < 0.740 < 0.829	0.480	0.52:1.31
WD	0.667 < 0.731 < 0.795	0.462	0.50:1.33	0.751 < 0.826 < 0.900	0.651	0.58:1.50
WD_HG	0.666 < 0.730 < 0.794	0.460	0.55:1.34	0.792 < 0.860 < 0.927	0.719	0.64:1.48
BA_HG	0.658 < 0.725 < 0.792	0.450	0.54:1.43	0.764 < 0.838 < 0.912	0.676	0.51:1.45
BA_WD	0.708 < 0.769 < 0.830	0.539	0.59:1.40	0.813 < 0.876 < 0.939	0.752	0.62:1.60
WD_HG_BA	0.717 < 0.777 < 0.837	0.555	0.57:1.43	0.829 < 0.889 < 0.949	0.779	0.63:1.57

Table 11 AUCs, Gini-coefficients and Jouden-Indices for best cutoffs (Co) for symmetrical classification (specificity = sensitivity) for the logistic models applied to both patient cohorts.

Fig 7 ROC-curves for $P(BA, WD, HG)$, $P(BA, WD)$, and $P(WD)$. At the top we see the curves for the basic 351-patient cohort and at the bottom the corresponding curves for the 120 test patients.

Table 11 list Jouden-indices, AUCs (95%-confidence intervals) and Gini-coefficients for both cohorts and for all significant models. The Gini-coefficients and the AUCs show the same qualitative behavior for both cohorts though the values for the 351-group are systematically smaller. The Jouden-Indices for multivariate models suggest larger cutoffs than $P = 0.5$ for optimum model quality. Obviously the classification cut-off $P = 0.5$ used so far favors true positives rather than true negative outcomes. In the case on bivariate models smaller cutoffs are sometimes better.

Predictors	351 Patients			120 Test-Patients		
	AUC	Gini	Co:Jouden	AUC	Gini	Co:Jouden
BA	0.636 < 0.704 < 0.773	0.409	0.46:0.40	0.729 < 0.809 < 0.889	0.618	0.46:0.45
HG	0.560 < 0.632 < 0.705	0.265	0.48:0.21	0.651 < 0.740 < 0.829	0.480	0.52:0.31
WD	0.667 < 0.731 < 0.795	0.462	0.50:0.33	0.751 < 0.826 < 0.900	0.651	0.58:0.50
WD_HG	0.666 < 0.730 < 0.794	0.460	0.55:0.34	0.792 < 0.860 < 0.927	0.719	0.64:0.48
BA_HG	0.658 < 0.725 < 0.792	0.450	0.54:0.43	0.764 < 0.838 < 0.912	0.676	0.51:0.45
BA_WD	0.708 < 0.769 < 0.830	0.539	0.59:0.40	0.813 < 0.876 < 0.939	0.752	0.62:0.60
WD_HG_BA	0.717 < 0.777 < 0.837	0.555	0.57:0.43	0.829 < 0.889 < 0.949	0.779	0.63:0.57

Table 11 AUCs, Gini-coefficients and Jouden-Indices for best cutoffs (Co) for symmetrical classification (specificity = sensitivity) for the logistic models applied to both patient cohorts.

Discussion

In this work we investigated the impact of early in-patient rehabilitation along with medical treatment in a specialized acute geriatric ward. The aim of this paper was to study and predict the improvements of the patient's condition from admission to discharge using standardized screening parameters taken at admission. As in acute geriatric care typically outcome measures [5] combine ADLs with mobility we use in this study BD $\geq$ 60 for ADLs at discharge, which is the approved level of independence for acceptance in a further rehabilitation unit. The outcome target for mobility is a successful but dichotomous Timed-Up-and-Go-Test (TUGd = possible). TUG is not only an indicator for mobility but correlates also with cognition [22, 25]. Hence TUG-performance and Barthel score which together form the endpoint correspond to a reasonable level of physical functioning and independence after treatment. With prognostic models we can quantify a patient's potential to reach this endpoint. We started our analysis with 580 randomly chosen patients and investigated predictors for BD $\geq$ 60 and TUGd separately and then in combination as the endpoint.

Descriptive statistics showed that the Barthel-score in was significantly higher at discharge than at admission, with a strong effect size in all ICD-10 groups. Patients starting the treatment with a BA between 20 and 55 points have the largest improvement: Their BD was more than 20 points higher, which corresponds to a substantial reduction of the assistance needed after discharge. If 20 points is the average improvement, we see that 40 Barthel-points at admission indicate that a patient may leave the hospital with a mild ADL-impairment (BD $\geq$ 60).

In order to find predictors for BS-improvement we use linear regression models and classify the relevance of the significant predictors by effect sizes. As expected, the Barthel score at admission (BA) has the strongest impact. This is in accordance with the literature [26, 27]. HG, WD and MMSE follow in descending order. GDS-15 and TUG-time are weak predictors, however there is correlation between TUG and the ADLs [21]. Note that age had only a small influence on the BS-improvement (BD-BA). Weight and duration of treatment were found to be non-significant for BD. These findings are in accordance with a meta-analysis [31], which showed that ADL-improvement in acute geriatric care is dependent neither on age nor the days of hospitalization.

TUG-testing is a problem for geriatric patients because many cannot stand up und this results in a large floor effect. The analysis of the TUG-testing showed that only 194 (33%) of the patients could perform the test at admission and 406 (70%) at discharge. The number of patients with a TUG time under the 12 seconds limit [24] for normal mobility increased from 6 at admission to 47 at discharge. These small numbers indicate, that TUG-times are not suitable as criteria for mobility of acute geriatric patients. Large floor-effects are seen in the above cited meta-analysis [31] as well, where the data of almost 154.000 patients were analyzed, but only 80.700 data sets for the timed TUG were available. TUG-times are supposed to be indicators for fall risks, but as it was shown in [12] a dichotomous TUG has the same predictive quality as the timed TUG. Moreover, typical geriatric patients will not achieve normal functional mobility after treatment, with TUG-times at admission of around 30s and an averaged improvement of 10s [31]. These are approximately the same times found for our patients. The mean TUG time improved significantly from 26.7s to 17.9s (n = 194). If the level of assistance needed after discharge is of interest, the dichotomous TUGd is sufficient. Less assistance is required, if a patient can stand up and walk even if the TUG time is 20s or higher.

The important predictors for TUGd with a strong effect size are the same as found in the BD-analysis: BA, WD, HG. MMSE and HG loose some importance and WD seems to be more relevant for TUGd than HG, which is obvious, since TUG and WD are mobility indicators [16, 17, 24]. Age is significant but with a small effect size. Weight, GDS and hospitalization in days have non-significant effects.

If the endpoint, i.e. the combination of ADLs and mobility (BD $\geq$ 60 and TUGd possible) is investigated, we considered only those patients with BA < 60 to avoid biasing (351 of the 580 patients). Due to the smaller BAs, the metric scores at admission of these 351 patients show a systematic but understandable negative offset as compared to the full cohort (Tab. 1 and Tab. 5). However, the effect on TUG-performance is drastic. Only 50 patients (14%) with BA < 60 performed the TUG at admission instead of 33% in the full group. In total 184 of the 351 patients (52% instead of 70% of the full group) could do the TUG at discharge. This shows that lower BAs result in lower TUGd performance.

Those, who could do the TUGd at discharge have significantly higher Barthel-improvements (Δ BS) of 27.2 as compared to those who failed: 15.4. However it should be noted that 20 patients who could perform the TUG at discharge, didn't reach a BD $\geq$ 60, whereas 58 with BD $\geq$ 60 couldn't do the TUGd. Obviously, TUGd is harder to achieve than BD $\geq$ 60. This confirms the expected relation [21] between ADLs and TUG, because both measures reflect the mobility-status of a patient. BS is widely accepted [26] as the fundamental measure for functionality in daily living. However, Barthel-testing is less strict in quantifying the transfer capability and walking.

For reaching the endpoint (BD $\geq$ 60 and TUGd possible) BA ($f = 0.653$) and WD ($f = 0.443$) are the strongest predictors, but with smaller effect sizes as found for the models for TUGd (see Tab. 4). HG ($f = 0.313$) and MMSE ($f = 0.305$) only show a medium effect strength. Age, weight, GDS and hospitalization in days have a non-significant effect ($p > 0.1$).

This ranking of the effect sizes maybe interpreted as follows. BA was shown to be relevant for BD. Walking is a valid and reliable quantifier for physical function [16] and a part of the TUG-test. HG is an indicator for overall static physical strength [27,28] and hence a strong HG favours higher BD, (see Tab. 3), but is less relevant for the performance of the TUG. HG tests only static strength whereas the TUG performance has multifactorial components and a dynamic aspect. MMSE keeps the forth position in the effect size ranking, already found for the independent regressions for BD and TUGd. This may be either related to moderate impact on physical function, or, more likely, to the small spread of MMSE-scores found in our data. Our patients typically have no relevant impairments in cognition ($M = 24.3$ $SD = \pm 5.0$, Median 26.0) and hence small differences in MMSE-scores can only produce limited differences in the outcome.

Odds-Ratios for reaching the endpoint support the ranking of metric predictors by effect sizes. Significant higher Odds for reaching the endpoint were found for BA ≥ 40 , WD > 20m, MMSE ≥ 24 , and HG ≥ 25 kPa. In case of the surgical patients only conservative treatment is significantly favorable as compared to an operation, and patients admitted from an internal ward have higher Odds than those from a surgical ward. The effect sizes for BA and WD are medium sized, all others Odds are either not significant or exhibit weak effects. Again BA and WD are most important, followed by MMSE and HG. Note, that age < 85 and male/female yield non-significant Odds.

However, as only elderly patients are admitted and age is a marker for functional decline, age is expected to correlate with our strong predictors and hence be confounding. However, bivariate linear regression for BA, WD, MMSE with age as predictor result in non-significant models. HG is the only predictor that exhibits a significant correlation with age as reported already in [6] and declines by 0.29 kPa/(year of age), which is of no clinical relevance. Hence age is no relevant confounder for the dominant predictors. Even the oldest benefit from the comprehensive treatment and improve their ADLs and mobility significantly as already noted in [31].

Logistic regression models with the most important predictors as variables yield probabilities (P) for reaching the endpoint. In this way, patients with the highest rehabilitation potential maybe identified.

Prognosis in medical decision-making and the prognostic value of geriatric assessments was investigated by several authors [7, 8, 9, 10, 29, 32]. The main focus was survival rates after critical clinical conditions, fall risks, hospitalization and mortality, but to our knowledge, no prognostic models for the outcome of acute geriatric care was published so far. Here we derived probabilities P , that a patient in an acute geriatric ward with a certain set of assessment scores in BA, HG and WD at admission reaches a well-defined endpoint in ADLs and mobility at discharge.

The predictive power of seven multivariate and bivariate models was analyzed: $P(\text{BA, HG, WD})$, $P(\text{BA, WD})$, $P(\text{BA, HG})$ and $P(\text{HG, WD})$, $P(\text{BA})$, $P(\text{WD})$, $P(\text{HG})$. We compared sensitivities, specificities and the correct classification rate with $P = 0.5$ as the classification cut-off for the basic set of 351 patients ($\text{BA} < 60$) and for a different set of 120 test patients with $\text{BA} < 60$, not used for model derivation. The classification rates for the test-set as well as for the full cohort of 351 patients are similar but the classification works slightly better for the test-set. As the 95% confidential intervals of the AUCs overlap this may be a random effect. The full model $P(\text{BA, WD, HG})$ yields the best correct classification rates, followed by $P(\text{BA, WD})$ as a two-predictor model. $P(\text{BA})$ and $P(\text{WD})$ are the best bivariate models. However, the $P(\text{WD, HG})$ based on simple and quick assessments yields a sufficient measure to explore the potential of a patient, e.g. by using the contour plot (Fig. 5).

ROC-curves are frequently used [7, 29, 32] for a systematic evaluation of model quality. We see that for both cohorts the three predictor-model has the largest area under the curve and hence gives the best classification. The model with BA and WD is confirmed to be second best and from the one-dimensional models either $P(\text{BA})$ or $P(\text{WD})$ are useful. The standard classification cut-off $P = 0.5$ results in suboptimal classification results since positive rather than true negative outcomes are preferred. The best cutoff-point for symmetrical sensitivity and specificity, i.e. the best Jouden-index, is found for larger cutoffs than 0.5 for multivariate models and close to 0.5 for bivariate models. Hence bivariate-models seem to be more specific and less sensitive as compared to multivariate models.

A comparison of model accuracy in terms of AUCs to other prognostic models in geriatric research imply that our models for predicting the level of independence and physical functioning show an outstanding predictive quality. Predictions for future falls on the basis of TUG-times or other mobility related criteria e.g. have AUCs in range 0.51 to 0.67 [29,32], whereas our models show AUCs from 0.7 to 0.89. Note that 0.5 is the AUC of a random classifier and 1.0 is the maximum value for the perfect model. Therefore, prognosis of future falls seems to be more complicated, as the limited predictive quality of the models in [29, 32] show. However, unexpected hospitalization within a certain time-period and 3- or 5-years mortality can be predicted with similar prognostic strength [7], since similar AUCs are reported as in our investigation.

Limitations

In total, we investigated the data of randomly chosen 700 patients. Bias was minimized by using bootstrap resampling methods to derive 95% confidential-intervals and by considering patients from a 10-years-time span. The main limitation is the single centric character of the study. Moreover, the hospital is located in a rural area with an almost mono-ethnic population with a high fraction of patients working in farming and industry. Hence the physical background is different from patients in metropolitan areas. Data from other geriatric wards would be highly welcome to test our approach.

Our prognostic models address the potential for independent living but whether a patient in fact was able to return to home depends on his or her social situation, which is not under the control of the hospital. Since our study is retrospective there is no possibility for a follow up. A longitudinal study would therefore be of great interest. These questions will be addressed in future investigations.

Conclusions

In this study, we showed that early rehabilitation in acute geriatric care improves significantly ADL-capability and physical function. Whether the improvements are sufficient for a discharge to domesticity can be predicted by reliable prognostic models, which explore the potential of a patient already at admission using standardized screening parameters for the functional status.

A BD $\geq$ 60 and the ability to perform the TUGd is a valid endpoint and useful criterion to quantify the level of physical functioning and independence at discharge. TUGd and BD are complementary and predictive measures for the amount of assistance needed by a patient after returning to domesticity, because this combination addresses a wide range of competences as ADLs, cognition and motoric capabilities. Patients reaching the endpoint have the potential to be discharged to home or to live with reduced assistance in institutionalized care or in nursing homes.

The predictive quality of screening tests at admission was analyzed using linear and logistic regressions for BD, then TUGd and finally for their combination defining the endpoint. Significant predictors with strong effect size are the assessments of BA, WD, HG and to a minor extent MMSE. That means, the functional status at admission has the highest impact on the status at discharge. Age, the kind of disorder (internal or surgery), gender or the duration of hospitalization show no relevance. In fact, all patients benefit from the treatment of their disorders and the comprehensive rehabilitative training in the acute geriatric ward. These results correspond to the findings of surveys including more than 100.000 geriatric patients and this implies that our cohort of 580 patients represents a typical cross section.

Once the important predictors are known, we can calculate the probability that a patient reaches the endpoint using prognostic models. These models determine the potential of a patient for a discharge to home on the basis of the functional status measured at admission, quantified with BA, WD and HG. The use of such models enables discharge planning at an early stage of treatment in hospital. Different combinations of predictors in the logistic models yield correct classification in up to 80% of the cases. The model with all three important predictors (BA, HG, WD) is the most accurate, however (BA, WD) is almost as reliable. For practical purposes the model not relying on a time-consuming ADL-assessment (HG, WD) is also efficient to assess the potential of a patient. Bivariate-models (BA) or (WD) are less sensitive than their multivariate counterparts but have a higher specificity. AUCs verify that these models have a profound predictive quality.

It was verified that the models once established for a representative cohort, are useful to predict the outcome of the geriatric treatment for other patients as well which were not considered for model development. Indeed, the prognostic models worked surprisingly well in the test-cohort. This was validated by comparing ROC-curves, Jouden-indices and Gini-coefficients for the basic cohort with the results for the test patients.

As a basic rule, which can be applied directly in clinical practice, we find that a substantial BA-score of around 40 is the best starting point for the comprehensive treatment in acute geriatric care, which should be backed up by walking distance of more than 25 m. Patients with these competences at admission are very likely to reach a basic level of independence at discharge. We emphasize that some parameters of the patient's status at admission, which seem to be important at first glance, lack any predictive quality. In particular, these are clinical condition, the medical treatment before admission, length of hospitalization, weight, age or gender. They are non-predictive for the outcome of early rehabilitative care if the potential for increased mobility and independence or even a discharge to home is considered. Hence according to our results all kinds of geriatric patients benefit from acute rehabilitative care of the elderly.

Literature

1. Kosse NM, Dutmer AL, Dasenbrock L, Bauer JM, Lamothe CJC. Effectiveness and feasibility of early physical rehabilitation programs for geriatric hospitalized patients: a systematic review. *BMC Geriatr* 2013; **13**, 107. doi.org/10.1186/1471-2318-13-107
2. OECD, Health at a Glance 2019: OECD Indicators, OECD Publishing, Paris 2019; doi.org/10.1787/4dd50c09-en
3. Fox MT, Sidani S, Persaud M, Maimets I, Brooks D, Brien K. Acute care for elders, components of acute geriatric unit care: A systematic descriptive review. *J Am Geriatr Soc* 2013, **61**: 939–946. doi.org/10.1111/jgs.12282
4. Baztán JJ, Suárez-García FM, López-Arrieta J, Rodríguez-Mañas L, Rodríguez-Artalejo F. Effectiveness of acute geriatric units on functional decline, living at home, and case fatality among older patients admitted to hospital for acute medical disorders: meta-analysis *BMJ* 2009, **338** b50. doi.org/10.1136/bmj.b50
5. Heldmann P, Werner C, Belala N, Bauer JM, Hauer K. Early inpatient rehabilitation for acutely hospitalized older patients: a systematic review of outcome measures. *BMC Geriatr*. 2019; **19**, 189. doi.org/10.1186/s12877-019-1201-4
6. Rijk MJ, Roos P, Deckx L, van den Akker M, Buntinx F. Prognostic Value of handgrip strength in people aged 60 years and older: A systematic review and meta-analysis, *Geriatr. Gerontol Int* 2016, **16**: 5-20. doi.org/10.1111/ggi.12508
7. Zucchelli A, Vetrano DL, Grande G, Calderón-Larrañaga A, Fratiglioni L, Marengoni A, Rizzuto D. Comparing the prognostic value of geriatric health indicators: a population-based study. *BMC Med* 2019, **17**: 185. doi.org/10.1186/s12916-019-1418-2
8. Gill TM. The central role of prognosis in clinical decision making. *JAMA*. 2012; **11**,307(2):199-200. doi.org/10.1001/jama.2011.1992
9. Hamaker ME, Vos AG, Smorenburg CH, de Rooij SE, van Munster BC. The value of geriatric assessments in predicting treatment tolerance and all-cause mortality in older patients with cancer. *Oncologist* 2012, **17**: 1439-49. doi.org/10.1634/theoncologist.2012-0186.
10. Selacovic I, Dubljanin-Rapopovic E, Markovic-Denic L, Marusic V, Circovic A, Kadija M, Tomanovic-Vujadinovic, Tulic G. Can early Assessment of Hand Grip Strength in older Hip Fracture Patients predict Functional Outcome. *PLOS ONE* 2019; **14**(8): e021322 doi.org/10.1371/journal.pone.0213223
11. Podsiadlo D, Richardson S. The Timed Up and Go. A test of basic functional mobility for frail elderly persons. *J Am Geriatr* 1991, **39**:142–148. doi.org/10.1111/j.1532-5415.1991.tb01616.x
12. Large J, Gan N, Basic D, Jennings N (2006) Using the Timed Up and Go Test to stratify elderly inpatients at risk of falls. *Clinical Rehabilitation* 2006, **20**: 421-428. doi.org/10.1191/0269215506cr959oa
13. Steiber N. Strong or Weak Handgrip? Normative Reference Values for the German Population across the Life Course Stratified by Sex, Age, and Body Height. *PLoS ONE* 2016, **11**(10): doi.org/10.1371/journal.pone.0163917
14. Syddall H, Cooper C, Martin F, Briggs R, Sayer AA. Is grip strength a useful single marker of frailty. *Age and Aging* 2003, **32**: 650-656. doi.org/10.1093/ageing/afg111
15. Neumann S, Kwisda S, Krettek C, Gaulke R. Comparison of the Grip Strength using the Martin- Vigorimeter and the JAMAR- Dynamometer: Establishment of Normal Values. *In Vivo* 2017, **31**: 917-924. doi.org/10.21873/invivo.11147

16. Imms FJ, Edholm OG. Studies of gait and mobility in the elderly. *Age and Ageing* 1981, **10**: 147-156. doi.org/10.1093/ageing/10.3.147
17. Harada ND, Chiu V, Stewart AL. Mobility-related function in older adults: assessment with a 6-minute walk test. *Arch Phys Med Rehabil.* 1999, **80**:837-41. [doi.org/10.1016/s0003-9993\(99\)90236-8](https://doi.org/10.1016/s0003-9993(99)90236-8).
18. Folstein M, Folstein S, Mc Hugh P. Mini-Mental State – A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975, **12**: 189-198. [doi.org/10.1016/0022-3956\(75\)90026-6](https://doi.org/10.1016/0022-3956(75)90026-6)
19. Brodaty H, Moore C. The Clock Drawing Test for dementia of the Alzheimer's type: A comparison of three scoring methods in a memory disorders clinic. *Int J Geriatr Psychiatry.* 1997, **12**: 619-627
20. Shulman K I, Gold D P, Cohen C A, Zuccherro C A. Clock-Drawing and Dementia in the Community: A Longitudinal Study. *Int. J. Geriatric Psychiatry.* 1993, **8**: 487-496. doi.org/10.1002/gps.930080606
21. Allgaier AK, Kramer D, Mergl R, Fejtkova S, Hegerl U (2011) Validität der Geriatrischen Depressionsskala bei Altenheimbewohnern: Vergleich von GDS-15, GDS-8 und GDS-4. *Psychiatr Praxis.* 2011, **38**: 280-286. doi.org/10.1055/s-0030-1266105
22. Van Patten R, Lee EE, Graham SA, Depp CA, Kim HC, Jeste DV, Twamley EW. The Utility of the Timed Up-and-Go Test in Predicting Cognitive Performance: A Cross-Sectional Study of Independent Living Adults in a Retirement Community. *Journal of Applied Gerontology.* 2020, **39**:1163-1168. doi.org/10.1177/0733464819872636
23. Field A. *Discovering statistics using IBM SPSS Statistics 4th Edition* London Sage Ltd.; 2013
24. Bischoff HA, Stähelin HB, Monsch AU, Iversen MD, Weyh A, von Dechend M, Akos R, Conzelmann M, Dick W, Theiler R. Identifying a cut-off point for normal mobility: a comparison of the timed 'up and go' test in community-dwelling and institutionalised elderly women. *Age Ageing.* 2003, **32**:315-20. doi.org/10.1093/ageing/32.3.315
25. Ansai HJ, de Andrade LP, Nagakawa TH, Vale FAC, Caetano MJD, Lord SR, Rebelatto JR. Cognitive Correlates of Timed Up and Go Subtasks in Older People With Preserved Cognition, Mild Cognitive Impairment, and Alzheimer's Disease. *Am J Phys Med Rehabil.* 2017, **96**:700-705. doi.org/10.1097/PHM.0000000000000722
26. Burckhardt H, Burger M. Ergebnisse und Prädiktoren der geriatrischen Frührehabilitation im Akutkrankenhaus. *Z Gerontol Geriatr.* 2012, **45**:138–145. doi.org/10.1007/s00391-011-0203-4
27. Cohen ME, Marino RJ. The tools of disability outcomes research functional status measures. *Arch Phys Med Rehabil.* 2000, **81**: 21–29. doi.org/10.1053/apmr.2000.20620
28. Syddall H, Cooper C, Martin F, Briggs R, Sayer AA. Is grip strength a useful single marker of frailty? *Age and Aging* 2003, **32**: 650-656. doi.org/10.1093/ageing/afg111
29. Kojima G, Masud T, Dendrick D, Morris R, Gawler S, Treml J. Does the timed up and go test predict future falls among British community-dwelling older people? Prospective cohort study nested within a randomized controlled trial. *BMC Geriatrics* (2015), **15**, 38-44, [doi.10.1186/s12877-015-0039-7](https://doi.org/10.1186/s12877-015-0039-7)
30. Fawcett T. Introduction to ROC Analysis, *Pattern Recognition Letters* (2006), **27**: 861-874, doi.org/10.1016/j.patrec.2005.10.
31. Kwetkat A, Lehmann T, Wittrich A. Geriatrische Frührehabilitation. *Z Gerontol Geriatr* (2014) **47**, 372–378 [doi 10.1007/s00391-014-0660-7](https://doi.org/10.1007/s00391-014-0660-7)
32. Gates S, Smith LA, Fisher JD, Lamb SE. Systematic review of accuracy of screening instruments for predicting fall risk among independently living older adults. *J.*

Availability of data and materials

Data are available on request to the author B. Hoppe (bernhard.hoppe@h-da.de)

Ethics approval and Consent to participate

This study was performed in accordance with the Declaration of Helsinki. Written informed consent to the use of their anonymized health data obtained during their stay in the hospital was provided from each participant during the standard procedures at admission. This has been approved by the ethics committee of the Krankenhaus Korbach, Korbach, Hesse, Germany. The study was retrospectively registered on 2nd July 2019 by the ethic commission of the hospital und filed under registration number (MG1/569/770/2019).

Competing interests

The authors confirm that there are no competing interests.

Authors' contributions

AH designed the study and collected the patients' data. BH analyzed and interpreted the data using statistical methods. AH analyzed the clinical impact of the analysis. Both authors contributed on equal shares in writing the manuscript and have read and approved the final manuscript.

Acknowledgements

We thank Krankenhaus Korbach and in particular the clinical director Dr. A. W. Schneider for the support and encouragement making this study feasible.

Figures

Figure 1

Barthel Scores at admission (BA) and before discharge (BD) and BS-improvement with standard deviations indicated by black lines

Figure 2

Linear regression models for BD vs. predictor ranges in %: BA (solid line), MMSE (dashed line), HG (dash dotted) and WD (dotted)

Figure 3

Odds-ratios of predictors for reaching the endpoint calculated for 351 patients with BA < 60. Lines indicate the 95% interval, filled dots are ratios with a significant difference in chances for both groups.

Figure 4

Effect sizes of significant predictors in bivariate regressions for BD, TUGd (580 patients) and reaching the endpoint (351 patients with BA < 60 at admission) as outcome variables.

Figure 5

2D-plot for $P(BA, WD)$

Figure 6

Contour plot for $P(WD, HG)$ with contour lines with a 0.1-spacing. Red dashed line is the $P = 0.5$ line, which is used for classification. The intersection of $WD = 40m$ and $HG = 10kPa$ determines the probability to reach the endpoint.

Figure 7

ROC-curves for $P(\text{BA}, \text{WD}, \text{HG})$, $P(\text{BA}, \text{WD})$, and $P(\text{WD})$. At the top we see the curves for the basic 351-patient cohort and at the bottom the corresponding curves for the 120 test patients.