



Published in final edited form as:

*J Geriatr Oncol.* 2019 January ; 10(1): 120–125. doi:10.1016/j.jgo.2018.06.012.

## Gait Speed and Survival of Older Surgical Patient with Cancer: Prediction after Machine Learning

Keyvan Sasani<sup>1</sup>, Helen N. Catanese<sup>1</sup>, Alireza Ghods<sup>1</sup>, Seyed Ali Rokni<sup>1</sup>, Hassan Ghassemzadeh<sup>1</sup>, Robert J. Downey<sup>3</sup>, Armin Shahrokni<sup>2</sup>

<sup>1</sup>School of Electrical Engineering and Computer Science, Washington State University, Pullman, Washington, 99164, USA

<sup>2</sup>Department of Medicine / Geriatrics, Memorial Sloan Kettering Cancer Center, NY, NY, 10065, USA

<sup>3</sup>Department of Surgery, Memorial Sloan Kettering Cancer Center, NY, NY, 10065, USA

### Abstract

**Purpose:** Gait speed in older patients with cancer is associated with mortality risk. One approach to assess gait speed is with the ‘Timed Up and Go’ (TUG) test. We utilized machine learning algorithms to automatically predict the results of the TUG tests and its association with survival, using patient-generated responses.

**Methods:** A decision tree classifier was trained based on functional status data, obtained from preoperative geriatric assessment, and TUG test performance of older patients with cancer. The functional status data were used as input features to the decision tree, and the actual TUG data was used as ground truth labels. The decision tree was constructed to assign each patient to one of three categories: “*TUG <10 seconds*”, “*TUG = 10 seconds*”, and “*uncertain*.”

**Results:** In total, 1901 patients (49% women) with a mean age of 80 years were assessed. The most commonly performed operations were urologic, colorectal, and head and neck. The machine learning algorithm identified three features (cane/walker use, ability to walk outside, and ability to

---

**Corresponding author:** Armin Shahrokni, MD, MPH, Geriatric Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, Box 205, 1275 York Ave., New York, NY 10065, Phone: 646-888-3250, Fax: 646-888-3181, shahroka@mskcc.org.

Author Contributions:

Study concepts Armin Shahrokni, Hassan Ghassemzadeh

Study design Armin Shahrokni, Hassan Ghassemzadeh

Data acquisition Armin Shahrokni, Hassan Ghassemzadeh

Quality control of data and algorithms Armin Shahrokni, Hassan Ghassemzadeh, Keyvan Sasani, Helen N. Catanese, Seyed Ali Rokni

Data analysis and interpretation Armin Shahrokni, Hassan Ghassemzadeh, Keyvan Sasani, Helen N. Catanese, Seyed Ali Rokni

Statistical analysis Armin Shahrokni, Hassan Ghassemzadeh, Keyvan Sasani, Helen N. Catanese, Seyed Ali Rokni

Manuscript preparation Armin Shahrokni, Hassan Ghassemzadeh, Keyvan Sasani, Helen N. Catanese, Seyed Ali Rokni, Alireza Ghods

Manuscript editing Armin Shahrokni, Hassan Ghassemzadeh, Keyvan Sasani, Helen N. Catanese, Seyed Ali Rokni, Alireza Ghods

Manuscript review Armin Shahrokni, Hassan Ghassemzadeh, Keyvan Sasani, Helen N. Catanese, Seyed Ali Rokni, Alireza Ghods

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Presentation:** This study was not presented in any meeting.

Competing Interests

The authors declare that they have no competing interests.

perform housework), in predicting TUG results with the decision tree classifier. The overall accuracy, specificity, and sensitivity of the prediction were 78%, 90%, and 66%, respectively. Furthermore, survival rates in each predicted TUG category differed by approximately 1% from the survival rates obtained by categorizing the patients based on their actual TUG results.

**Conclusions:** Machine learning algorithms can accurately predict the gait speed of older patients with cancer, based on their response to questions addressing other aspects of functional status.

### Keywords

Machine learning; decision tree; predictive analytics; cancer; TUG; survival; gait speed

---

## BACKGROUND

The number of people aged 65 years or older is expected to rise from 43.1 million in 2012 to 83.7 million in 2050 [1]. Because the incidence of cancer increases with age, it is expected that more older patients will be diagnosed with cancer in the next decades. Among older patients with cancer, those who are frail are at the highest risk for developing adverse events, such as toxicity from chemotherapy, and complications from surgery [2]. Research shows that instead of age, a patient's level of fitness should inform cancer treatment decision making [3]. Geriatric oncology experts utilize a geriatric assessment (GA), to assess patients' fitness for cancer treatment. The GA is a multidimensional assessment of older patients with cancer that includes an assessment of functional status as well as other parameters such as weight loss or polypharmacy [3]. Functional status assessment of older adults with cancer in the GA relies on both subjective and objective assessments. For example, patients are asked to report whether they are independent in performing basic activities of daily living, such as grooming or feeding, and in instrumental activities of daily living, such as transportation or shopping [4,5]. To obtain an objective assessment of patients' functional status, geriatric oncology experts utilize tests such as the 'Timed Up and Go' test (TUG) [6]. In this test, a patient is instructed to get up from a chair without using their arms, walk ten feet, turn around, and return to the chair. The time to perform the TUG is measured in seconds [6]. In non-cancer settings, slower TUG speed has been associated with poor performance in cognition and memory tests, [7] and to predict risk for falls and fractures [8]. In a cancer setting, TUG has been used as a test within GA [9]. It is shown that older patients with cancer with slower TUG times are at twice the risk for developing postoperative complications compared to patients with faster times [10]. TUG speed has also been associated with early death with first line chemotherapy. In a study on 348 patients with a median age of 77 years, patients with slower TUG speed were 2.55 times more likely to die within six months of starting first line chemotherapy [11]. Despite these prognostic benefits, administering TUG presents challenges. Completing the TUG requires staff time, and the error rate between different observers could be substantial. In a study on 78 participants with an average age of 84 years, TUG was assessed by 20 physical therapists. TUG was checked in 3 sessions all within one week. One physical therapist assessed TUG on session 1 and 2, and then another physical therapist assessed TUG for the third session. The interquartile difference between the assessments in the first two sessions by the same

assessor, ranged from -2.1 to 3.2 seconds. The interquartile difference between assessments by the two assessors, ranged from -3.1 to 3.4 seconds. The variability increased as patients' TUG slowed. For example, the 95% confidence interval for a TUG of 20 seconds was 14.8 to 27 seconds, while the 95% confidence interval for a patient with a TUG of 30 seconds was 22.2 to 40.5 seconds [12]. Furthermore, in another study on 235 patients with advanced chronic obstructive pulmonary disease, congestive heart failure, or chronic renal failure, three TUG assessments were performed by one assessor. The absolute minimal detectable change was as high as 4.5 seconds, corresponding to as high as a 35% difference between the first and second trials [13]. In addition to these personnel-associated limitations, the TUG test needs to be conducted within a clinical facility and not remotely.

We have shown that some components of GA can be acquired remotely [14]. If these components can be shown to correlate with TUG times, performing the TUG test in clinics can be potentially avoided or at least reduced substantially.

The goal of this project was to utilize a machine learning approach to develop an algorithm based on components of the GA, other than TUG, to accurately predict which patients will have slower TUG times. Due to the demonstrated relationship between TUG and overall survival, we then compared the survival rate of patients based on their predicted TUG scores against the performed TUG results. If our algorithm predicts TUG accurately, then it can serve as a screening tool for TUG assessment. In this situation, TUG can be performed only on patients that may have slower TUG times as suggested by the machine learning algorithm.

## METHODS

The Memorial Sloan Kettering Cancer Center (MSKCC) Institutional Review Board authorized analysis of electronic medical records (EMR) of patients who presented at the Geriatrics Service Clinic for preoperative evaluation, and underwent surgery within a month at Memorial Hospital (MH). The data were collected between 1/2015 and 12/2016 and included patients who were 75 years or older. The patients completed a geriatric assessment through an electronic tool called the electronic Rapid Fitness Assessment or (eRFA) [14]. The geriatric assessment examines functional status of the patients through assessment of basic activities of daily living (bADL), and instrumental activities of daily living (iADL) [4,5].

The patients were asked to indicate whether they were unable to perform a task, needed some help to perform a task, or could perform a task without any help, for bADL and iADL activities. bADL assesses seven basic activities including bathing, dressing, grooming, feeding, walking inside the home, walking outside the home, and bladder-bowel control. iADL assesses eight functional activities including telephone use, doing laundry, shopping, preparing meals, doing housework, self-medication, handling money and finances, and transportation. In addition, the patients were asked to report whether they use a cane, walker, or both. The assessment also included questions about the presence and number of hours of home health aide assistance, level of social support [15], limitations in social activities [16],

depression [17], cognitive status[18,19], and two questions related to nutritional status and poly-pharmacy.

Geriatric nurses used TUG to determine each patient's gait speed. Geriatric nurses had an average of 10 years of working experience at MSKCC Geriatrics service, where TUG is performed in the geriatrics clinics as routine care. MSKCC Geriatricians reviewed and confirmed the TUG assessment performed by the geriatric nurses. The results of TUG were then stratified into three categories: “*less than ten seconds*”, “*ten to nineteen seconds*”, and “*twenty seconds or more*”, and were entered into patient's electronic medical records. Finally, we combined the results of the eRFA with other clinical evaluations to inform perioperative management [19], and construct our database.

## Data Collection

eRFA data was entered into an institutional database at the time of collection. Information from this dataset was extracted and integrated with socio-demographic characteristics (e.g. age, gender, marital status, educational status, and living condition), retrieved from the MH electronic medical record. Patient mortality was retrieved by linking the dataset to the Social Security Death Index.

## Data Cleaning

In the data cleaning phase, 134 patients were excluded from our data analysis because the TUG scores were not recorded at those entries. After this initial pre-processing, there were 4,936 missing items (3.8% of the total items) in the dataset. This amount of missing data was associated with only 5 variables. The missing values were imputed with the mean, median, or mode for that variable, depending on the characteristics of the variable [20]. Specifically, discrete unordered variables were filled via mode, continuous variables by mean, and any remaining variables were filled by medians.

## Machine Learning

The primary goal of this study was to build an interpretable model that allows a clinician to accurately predict a patient's TUG score with less than five GA questions. To achieve this goal, we focused on developing a machine learning approach that limits the number of questions to a maximum of three. Given answers to those questions, the machine learning algorithm predicts the patient's TUG score as one of the three categories (i.e., classes), including “*TUG less than ten seconds*”, “*TUG more than ten seconds*”, and “*uncertain*”. The “uncertain” category refers to cases where further testing will be needed.

We utilized the decision tree algorithm to produce an adaptive questionnaire. A decision tree algorithm is computationally simple and provides a human-interpretable output for prediction. In a decision tree, each internal (non-leaf) node is labelled with an input feature (i.e., a question). A node will branch to a child node based on the provided answer to the question associated with its parent node. Starting from a single node, called the “root”, the decision tree algorithm generates a tree structure where each non-leaf node is associated with a survey question, and each leaf in the tree is labelled with a class (i.e., the decision), or a probability distribution over the classes.

We utilized *scikit-learn* [20], a machine learning library for Python, for construction validation of the decision tree model. Constructing a decision tree began by choosing a question at each iteration that best split the remaining data samples [21]. We measured the informativeness of each question by calculating the homogeneity of the target variable within the subsets. Specifically, we calculated a score called Gini Impurity [22], that indicates how well each question can split patients with TUG scores of less or greater than ten seconds, into two homogenous subsets. The process of decision tree construction continued by selecting the most informative questions and adding them to the tree, until the model reached a specified complexity (e.g., tree depth), or it could categorize all the provided training data correctly [23,24].

In the process of building a predictive model using machine learning, it is important to provide the machine representative yet unbiased training data, to help the machine learn the concept (reasons of having different values for TUG) better. Because the original data distribution was biased to a specific class, we utilized an under-sampling method [25] to balance the classes. Under-sampling attempts to balance the number of positive and negative samples by eliminating some of the majority samples from the training dataset. For example, if the majority of our data have TUG scores of less than ten seconds, under-sampling methods only select a subset of data with TUG less ten seconds, to facilitate the learning procedure.

### Correlation Analysis

We performed correlation analysis to achieve two goals: 1) to show that the selected variables for building the proposed machine learning model had the strongest correlation with TUG, and 2) to provide researchers with the list of additional variables with strong correlations with the TUG, for future studies that will focus on external validation of our model. To identify the correlation strength between features in the eRFA and a patient's TUG score, we used the Kendall rank correlation coefficient (tau) [26]. This score provides a value that ranges from -1 to 1 where a larger absolute value corresponds to a stronger correlation, which is either positive or negative. The advantages of this approach are that it does not assume that the data are normally distributed or that the relation between variables is linear. Because the test assumes that features are ordered, we excluded non-ordinal features (e.g., marital status) from this correlation analysis.

### Association between Predicted and Actual TUG with Survival

We calculated patient survival rates within the two prediction groups, TUG > ten seconds and TUG < ten seconds, and compared the survival rates to those of patients with actual TUG > ten seconds and TUG < ten seconds, measured through TUG tests administered in clinics. The survival rates were calculated at the 60-, 180- and 360-day thresholds. Patients who underwent surgery and did not have adequate follow up time for each survival time analysis were excluded from this analysis.

## RESULTS

### Patients

In total, 1901 patients (median age of 80) were included in this analysis. Among all the subjects, 49% were women, 55.4% were married, and 48.4% were college graduates or had advanced degrees. The most common operations or procedures performed were urologic (24.7%), colorectal (9.7%), head and neck (9.5%), gynecologic (9.3%), hepatobiliary (8.3%), interventional radiologic (6.7%), thoracic (6.4%), and breast surgery (5.3%). The remaining surgical procedures were less than 5% of the sample. Among all the patients, 48.4% had outpatient procedures, and the rest required hospitalization. Table 1 shows a breakdown of characteristics of patients in each TUG category as well as those of all the patients.

### Decision Tree Algorithm

Figure 1 shows the learned decision tree algorithm for TUG score prediction. As the figure shows, it is possible to reach a leaf node by asking a patient, at most, three survey questions. The leaf nodes are associated with the three desired class labels, TUG > ten seconds, TUG < ten seconds, and uncertain. This figure also shows TUG score prediction accuracy. On average, the accuracy, specificity, and sensitivity of the developed decision tree were 78.04%, 90.28%, and 65.80%, respectively. In another word, of 1170 patients with actual TUG < 10 seconds, our model was able to predict TUG correctly in 913 patients (78%), incorrectly in 92 patients (7.9%), and was indeterminate in 165 patients (14.1%). On the other hand, out of 731 patients with actual TUG > 10 seconds, our model predicted TUG accurately in 473 patients (64.7%), incorrectly in 142 patients (19.4%), and was indeterminate in 116 patients (15.9%).

In general, our model could confidently predict a patient's condition 85% of the time. Our experiments show that the proposed under-sampling strategy improves the model accuracy from 65.2% to 80.8%.

### Comparison of Survival between Actual TUG and Algorithmic TUG

A t-test showed that there is a significant difference between the survival rates of patients classified as TUG<10 and those classified as either TUG > ten seconds, with a p-value <0.1. Similarly, a chi-squared test showed that the probability that the variation in survival rates of patients between predicted and actual TUG was due only to chance is above the 99% threshold.

To further assess the relationship between these TUG frailty predictions and patient outcomes, the survival rates of patients in various TUG classes were calculated. The 360-day survival rate included 1136 patients, the 180-day survival included 1619 patients, and the 60-day survival included 1882 patients. In Table 2, the 60-, 180- and 360-day survival rates of patients in different prediction groups are shown against the survival rates of the patients in the respective groups. The survival rates for actual and predicted TUG groups were comparable with about 1% variation, and patients with higher TUG scores had a lower survival rate, as expected.

## Correlation

**Analysis**—In addition to using machine learning to streamline the process of assessing patient TUG scores, we aimed to identify significant correlations between TUG scores and various demographic attributes available in a pre-op survey. This analysis helped us validate our decision tree model, since the strongest correlations are used in our model. This also gives us some intuition on which other variables are the candidates to add to the model, to get more accurate results. The ten strongest positive and negative correlations are shown in Figure 2. TUG scores were negatively correlated with all grooming and personal care functions in the dataset, in addition to use of movement assistive devices, home care, and social/activity interference. In particular, difficulty in performing the TUG test was associated with difficulty getting around both inside and outside, the ability to do housework, and the ability to prepare meals.

## DISCUSSION

The study's goal was to predict patient TUG results through machine learning algorithms. We were able to develop a very short algorithm to accurately predict different categories of TUG.

It is critical to determine patient functional activity, especially in older patients with cancer, to inform treatment decision making. Patients who are more active or fit are able to tolerate the cancer treatment more than patients who are inactive or frail [27]. In older patients with cancer, the process of assessing fitness is commonly through a geriatric assessment (GA) [3], which emphasizes functional status. Functional assessment can be done either via patient-reported instruments such as bADL and iADL, or via objective assessments such as TUG. It has been shown that TUG correlates with overall mortality in older patients with cancer receiving chemotherapy [11]. Slower TUG was associated with increased probability of surgical complications and one-year mortality in older surgical patients with cancer [28]. TUG can also predict future falls in community-dwelling older adults [29]. In our center, we use preoperative TUG to guide the need for requesting physical therapy evaluation in the postoperative period. Furthermore, because patients with slower TUG usually have other aging-related impairments and could be at risk for falls, as time permitted, we discussed home safety with the patients and families in the preoperative period. If following this conversation, more concerns arose, the outpatient geriatrics team communicated this to the surgical team, physical therapy team, and the case managers.

However, performing TUG is not without challenges. First, while the test takes about ten to twenty seconds per patient, performing such a test in fast-paced oncology clinics as routine care may not be feasible. As a result, there is a need to develop a short screening questionnaire or algorithm that can predict patient's TUG with high accuracy, sensitivity, and specificity. Our algorithm can be used to screen patients for the need to perform TUG prior to being seen by a clinician, limiting administration to only 15% of patients, because our algorithm was able to predict TUG in 85% of population. Our algorithm, consisting of only three questions - querying the patient's use of assistive device, ability to walk outside of home without any limitations, and ability to perform housekeeping tasks without any help - can predict TUG scores with 78% accuracy. To the best of our knowledge, our study is the

first to utilize a machine learning algorithm to develop a quick procedure for patient TUG performance prediction.

One of the limitations of this study is that it was conducted in a single institution as a retrospective study in a cohort of older patients with cancer at the time of preoperative evaluation. In addition, the proposed model could increase its reliability and certainty if the machine learning procedure had access to a larger surgical population data. Although appropriate methods were used to account for missing data, less missing data would be preferred. Future studies should be conducted on external validation of the model.

Different studies used different cutoff scores [30–32] for the TUG. However, a descriptive meta-analysis suggests that a TUG of nine to ten seconds be used as a cutoff for slow walking [33]. In our study, nurses categorized TUG scores into three groups. As a result, the exact TUG score was unknown. Future studies should focus on whether machines can predict the exact TUG. Analysis of the GA dataset showed that the data suffered from an imbalanced class distribution, where 62% of the patients had a TUG score of less than ten seconds, and less than 48% of the patients had a TUG score of greater than ten seconds. This imbalanced class distribution biases the algorithm towards predicting one of the classes. To address this problem, we utilized the previously discussed under-sampling method to balance the classes.

Despite this limitation, our study was performed on one of the largest datasets on older patients with cancer at the time of preoperative evaluation. All patients completed geriatric assessments as routine care, which eliminated healthy selection bias. Instead of reporting odds ratios, we have used a practical method to develop a step-by-step algorithm, which can be used in any fast-paced oncology clinic with limited resources.

Our motivation for choosing a decision tree as the machine learning algorithm is as follows:

- It is easy to understand and interpret which enables medical researchers to easily analyze and validate its results.
- A decision tree can handle both numerical and categorical questions included in the medical surveys [23].
- Decision trees are robust to outliers and noisy data [24].
- The structure of a decision tree is well suited to the requirement of reducing the number of questions.

These advantages make data preparation easier and more flexible in case of missing values. Based on the structure of the decision tree, more informative questions are closer to the root of the tree and less informative questions are closer to leaves of the tree.

## CONCLUSIONS

This study showed that a simple decision tree was able to predict patient gait speed with high accuracy. The proposed decision tree can be used to screen patients who need further functional assessment or intervention. Future prospective studies are needed to provide



external validation of our model, and assess its impact on health care processes, and outcomes of older patients with cancer.

## Acknowledgements

The authors would like to thank all patients and their families who completed the assessment. In addition, we appreciate the efforts by the geriatric nurses and nurse practitioners in assessing patients. We are grateful for the collaboration by our geriatricians.

**Financial support:** The project was supported, in part, by the Beatriz and Samuel Seaver Foundation, the Memorial Sloan Kettering Cancer and Aging Program, the NIH/NCI Cancer Center Support Grant P30 CA008748, and the National Science Foundation (NSF) under grants CNS-1566359 and CNS-1750679. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations.

## References

- [1]. Ortman JM, Velkoff VA, and Hogan H, An aging nation: the older population in the United States. United States Census Bureau, Economics and Statistics Administration, US Department of Commerce, 2014.
- [2]. Hamaker ME, Jonker JM, de Rooij SE, Vos AG, Smorenburg CH, and van Munster BC, "Frailty screening methods for predicting outcome of a comprehensive geriatric assessment in elderly patients with cancer: a systematic review," *Lancet Oncol*, vol. 13, pp. e437–e444, 2012. [PubMed: 23026829]
- [3]. Korc-Grodzicki B, Holmes HM, and Shahrokni A, "Geriatric assessment for oncologists," *Cancer Biol. Med*, vol. 12, pp. 261–274, 2015. [PubMed: 26779363]
- [4]. Katz S, "Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living.," *J. Am. Geriatr. Soc*, vol. 31, pp. 721–727, 1983. [PubMed: 6418786]
- [5]. Lawton M and Brody E, "Instrumental Activities of Daily Living Scale (IADL)," 1988.
- [6]. Podsiadlo D and Richardson S, "The timed" Up & Go": a test of basic functional mobility for frail elderly persons.," *J. Am. Geriatr. Soc*, vol. 39, pp. 142–148, 1991. [PubMed: 1991946]
- [7]. Donoghue OA, Horgan NF, Savva GM, Cronin H, O'regan C, and Kenny RA, Association between timed up-and-go and memory, executive function, and processing speed. *J Am Geriatr Soc*, vol. 60, pp. 1681–1686, 2012. [PubMed: 22985141]
- [8]. Zhu K, Devine A, Lewis JR, Dhaliwal SS, and Prince RL, "Timed Up and Go Test and Bone Mineral Density Measurement for Fracture Prediction.," *Arch. Intern. Med*, vol. 171, pp. 1655–1661, 2011. [PubMed: 21987195]
- [9]. Hurria A, Gupta S, Zauderer M, Zuckerman EL, Cohen HJ, Muss H, Rodin M, Panageas KS, Holland JC, Saltz L, and Kris MG, "Developing a cancer-specific geriatric assessment.," *Cancer*, vol. 104, pp. 1998–2005, 2005. [PubMed: 16206252]
- [10]. Huisman MG, Van Leeuwen BL, Ugolini G, Montroni I, Spiliotis J, Stabilini C, Carino ND, Farinella E, de Bock GH, and Audisio RA, "Timed Up & Go": A Screening Tool for Predicting 30- Day Morbidity in Onco-Geriatric Surgical Patients? A Multicenter Cohort Study.," *PLoS One*, vol. 9, p. e0086863, 2014.
- [11]. Soubeyran P, Fonck M, Blanc-Bisson C, Blanc JF, Ceccaldi J, Mertens C, Imbert Y, Cany L, Vogt L, and Dauba F, Andriamampionona J, "Predictors of early death risk in older patients treated with first-line chemotherapy for cancer.," *J. Clin. Oncol*, vol. 30, pp. 1829–34, 2012. [PubMed: 22508806]
- [12]. Nordin E, Rosendahl E, and Lundin-Olsson L, "Timed 'Up & Go' Test: reliability in older people dependent in activities of daily living—focus on cognitive state," *Phys. Ther*, vol. 86, no. 5, pp. 646–655, 2006. [PubMed: 16649889]
- [13]. Mesquita R, Janssen DJA, Wouters EFM, Schols JMGA, Pitta F, and Spruit MA, "Within-Day Test-Retest Reliability of the Timed Up & Go Test in Patients With Advanced Chronic Organ Failure," *Arch. Phys. Med. Rehabil*, vol. 94, no. 11, pp. 2131–2138, 11 2013. [PubMed: 23583345]

- [14]. Shahrokni A, Tin A, Downey RJ, and Strong V, "Electronic Rapid Fitness Assessment: A Novel Tool for Preoperative Evaluation of the Geriatric Oncology Patient," *J. Natl. Compr. Cancer Netw*, vol. 15, no. 2, pp. 172–179, 2017.
- [15]. Gjesfjeld CD, Greeno CG, and Kim KH, "A confirmatory factor analysis of an abbreviated social support instrument: The MOS-SSS," *Res. Soc. Work Pract*, vol. 18, no. 3, pp. 231–237, 2008.
- [16]. Stewart AL and Ware JE, *Measuring functioning and well-being: the medical outcomes study approach*. duke university Press, 1992.
- [17]. Almeida OP and Almeida SA, "Short versions of the geriatric depression scale: a study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV," *Int. J. Geriatr. Psychiatry*, vol. 14, no. 10, pp. 858–865, 1999. [PubMed: 10521885]
- [18]. Borson S, Scanlan JM, Chen P, and Ganguli M, "The Mini-Cog as a screen for dementia: validation in a population-based sample," *J. Am. Geriatr. Soc*, vol. 51, no. 10, pp. 1451–1454, 2003. [PubMed: 14511167]
- [19]. Borson S, Scanlan J, Brush M, Vitaliano P, and Dokmak A, "The Mini-Cog: a cognitive 'vital signs' measure for dementia screening in multi-lingual elderly," *Int. J. Geriatr. Psychiatry*, vol. 15, no. 11, pp. 1021–1027, 2000. [PubMed: 11113982]
- [20]. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, and Duchesnay É, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res*, vol. 12, no. Oct, pp. 2825–2830, 2011.
- [21]. Quinlan JR, "Induction of decision trees," *Mach. Learn.* 1, vol. 1, no. 1, pp. 81–106, 1986.
- [22]. Grabmeier JL and Lambe LA, "Decision trees for binary classification variables grow equally with the Gini impurity measure and Pearson's chi-square test," *Int. J. Bus. Intell. Data Min*, vol. 2, no. 2, pp. 213–226, 2007.
- [23]. Fayyad UM and Irani KB, "On the handling of continuous-valued attributes in decision tree generation.," *Mach. Learn*, vol. 8, no. 1, pp. 87–102, 1992.
- [24]. Friedl MA and Brodley CE, "Decision tree classification of land cover from remotely sensed data," *Remote Sens. Environ*, vol. 61, no. 3, pp. 399–409, 9 1997.
- [25]. Chawla NV, "C4. 5 and imbalanced data sets: investigating the effect of sampling method, probabilistic estimate, and decision tree structure," *Proc. ICML*, vol. 3, 2003.
- [26]. Kendall MG, "A New Measure of Rank Correlation," *Biometrika*, vol. 30, no. 1–2, pp. 81–93, 1938.
- [27]. Hurria A, Togawa K, Mohile SG, Owusu C, Klepin HD, Gross CP, Lichtman SM, Gajra A, Bhatia S, Katheria V, and et al., "Predicting chemotherapy toxicity in older adults with cancer: a prospective multicenter study," *J. Clin. Oncol*, vol. 29, no. 25, pp. 3457–3465, 2011. [PubMed: 21810685]
- [28]. Robinson TN, Wu DS, Sauaia A, Dunn CL, Stevens-Lapsley JE, Moss M, V Stiegmann G, Gajdos C, Cleveland JC Jr, and Inouye SK, "Slower walking speed forecasts increased postoperative morbidity and one-year mortality across surgical specialties," *Ann. Surg*, vol. 258, no. 4, p. 582, 2013. [PubMed: 23979272]
- [29]. Shumway-Cook A, Brauer S, and Woollacott M, "Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test," *Phys. Ther*, vol. 80, no. 9, pp. 896–903, 2000. [PubMed: 10960937]
- [30]. Bischoff HA, Stähelin HB, Monsch AU, Iversen MD, Weyh A, Von Dechend M, Akos R, Conzelmann M, Dick W, and 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*, vol. 32, no. 3, pp. 315–320, 2003. [PubMed: 12720619]
- [31]. Thrane G, Joakimsen RM, and Thornquist E, "The association between timed up and go test and history of falls: the Tromsø study," *BMC Geriatr*, vol. 7, no. 1, p. 1, 2007. [PubMed: 17222340]
- [32]. Nordin E, Lindelöf N, Rosendahl E, Jensen J, and Lundin-Olsson L, "Prognostic validity of the Timed Up-and-Go test, a modified Get-Up-and-Go test, staff's global judgement and fall history in evaluating fall risk in residential care facilities," *Age Ageing*, vol. 37, no. 4, pp. 442–448, 2008. [PubMed: 18515291]

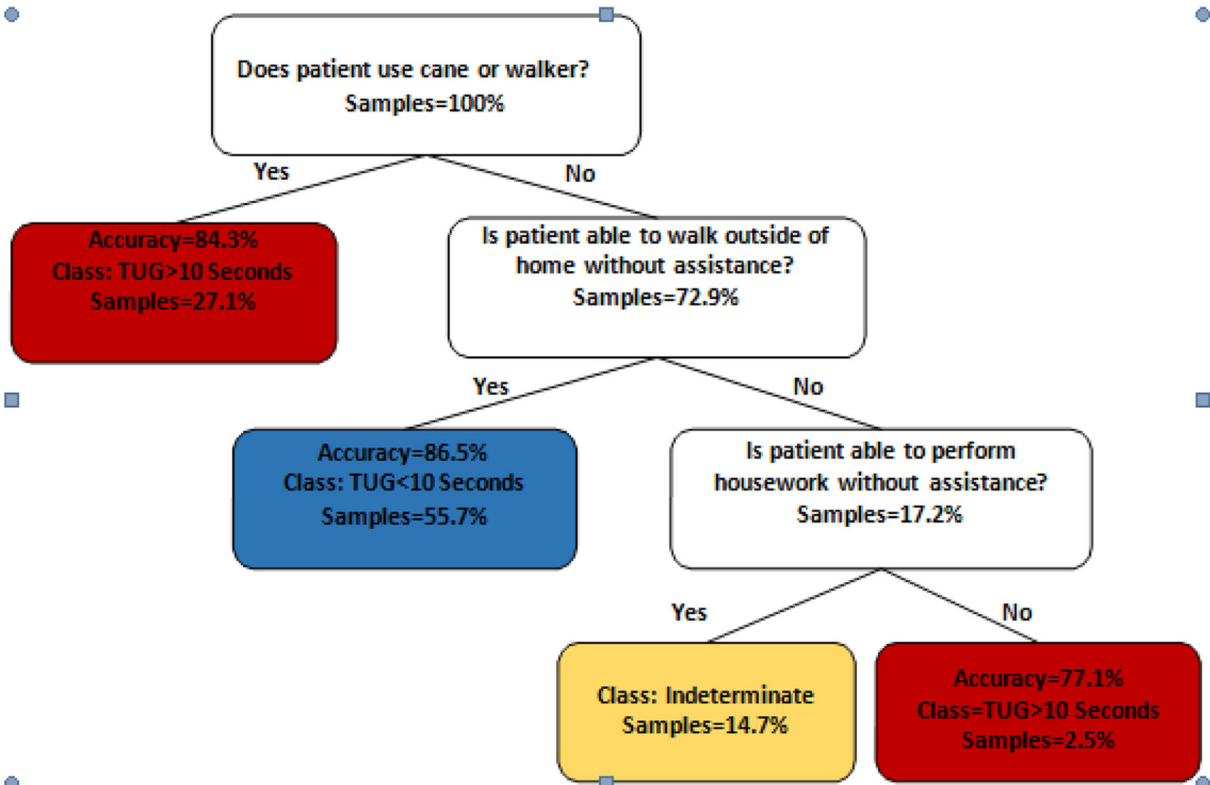
- [33]. Bohannon RW, "Reference Values for the Timed Up and Go Test," *J. Geriatr. Phys. Ther.*, vol. 29, no. 2, pp. 64–68, 2006. [PubMed: 16914068]

Author Manuscript

Author Manuscript

Author Manuscript

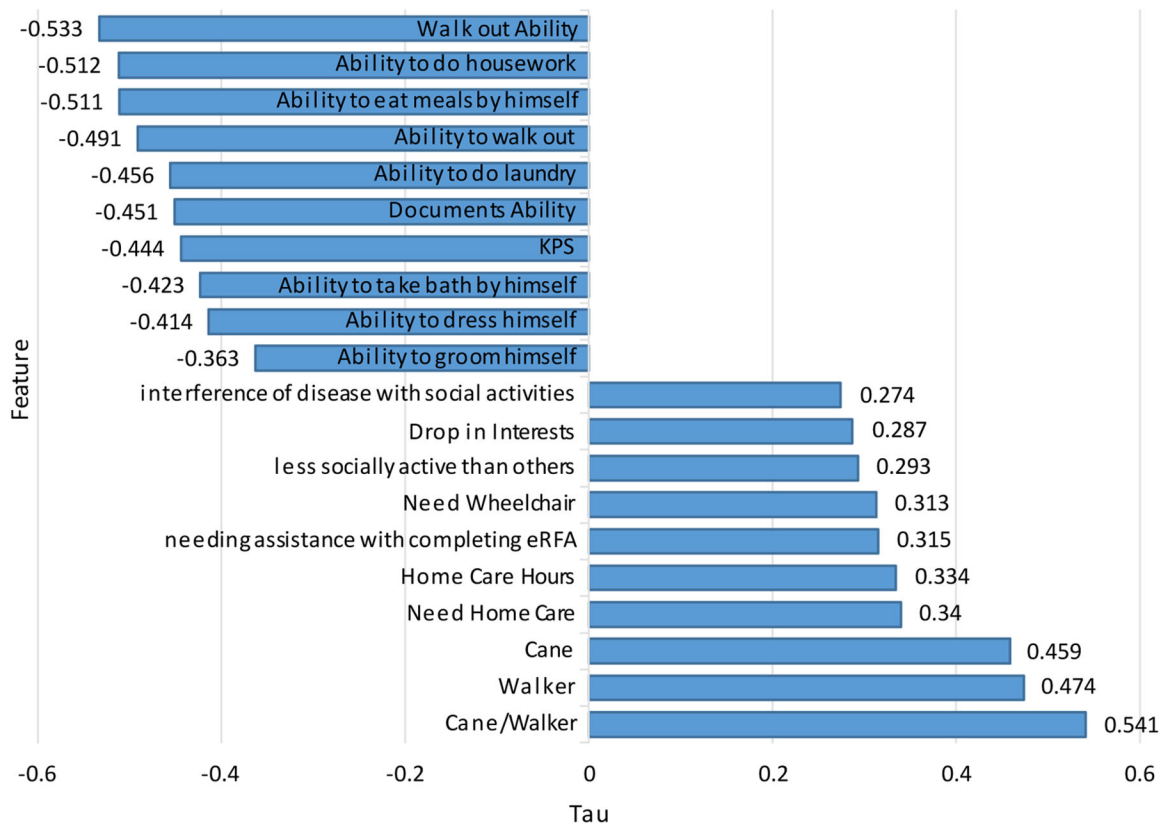
Author Manuscript



**Figure 1.**

The Timed Up and Go test (TUG) decision tree. Each leaf represents a classification of either TUG > ten seconds, TUG < ten seconds, or indeterminate. The accuracy of a given prediction, and samples percentage, from the original dataset which fall into each leaf are included as well. The model suggests that the get up and go frailty indicator is highly related to the fact that whether a patient uses an assistive device to walk or not. If a patient uses a cane or walker, his TUG score is > ten seconds with an accuracy of 84.3%. On the other hand, if the patient does not use a cane or walker to walk, it depends on his or her ability to walk outside the house. Then, if s/he is able to walk outside without assistance, with 86% accuracy, s/he has TUG less than ten seconds. However, if the patient's walking outside function is impaired, and cannot do daily housework, the TUG is > ten seconds, with accuracy more than 77%.

### Correlation with TUG



**Figure 2.** The Kendall Rank Correlation Coefficients between Time Up and Go test (TUG) and ten positive and negative correlations.

**Table 1.**

Patients' sociodemographic characteristics and geriatric assessment per TUG category and overall

	<b>TUG &lt; 10 seconds (1170, 61.5%)</b>	<b>TUG 10–19 seconds (493, 25.9%)</b>	<b>TUG 20 seconds (238, 12.5%)</b>	<b>Missing TUG (134, 6.6%)</b>	<b>P value</b>
<i>Age (mean/SD)</i>	79.45 (4.1)	81.51 (5.0)	82.62 (5.4)	80.47 (4.8)	<0.001
<i>Gender</i>					
- <i>Female</i>	509 (43.5%)	275 (55.8%)	147 (61.8%)	68 (50.7%)	<0.001
- <i>Male</i>	661(56.5%)	218 (44.2%)	91(38.2%)	66 (49.3%)	
<i>Married</i>	696 (59.6%)	244 (49.5%)	92 (38.7%)	70 (54.7%)	<0.001
<i>Living with family/partner</i>	826 (70.7%)	304 (61.8%)	135 (57%)	77 (59.7%)	<0.001
<i>College graduate or higher</i>	608 (52.4%)	198 (40.3%)	96 (40.7%)	64 (48.6%)	<0.001
<i>KPS</i>	90.12 (11.3)	78.53 (16.2)	62.65 (16.5)	82.05 (17.19)	<0.001
<i>bADL score</i>	13.28 (1.2)	11.54 (2.7)	8.44 (3.8)	11.92 (3.3)	<0.001
<i>iADL score</i>	14.99 (2.2)	12.39 (3.8)	8.6 (4.5)	12.98 (4.4)	<0.001
<i>Social support score</i>	16.52 (3.9)	16.39 (3.7)	16.66 (3.4)	15.03(4.4)	0.66
<i>Social activity limitation score</i>	7.57 (2.3)	9.08 (2.6)	10.59 (2.8)	8.70 (2.9)	<0.001
<i>Depression score</i>	0.75 (0.9)	1.21 (1.2)	1.85 (1.2)	1.12 (1.0)	<0.001
<i>Mini-Cog score</i>	4.1(1.2)	3.63 (1.4)	3.01 (1.7)	3.45 (1.7)	<0.001
<i>Type of surgery</i>					
- <i>Same day</i>	544 (48%)	225 (50.3%)	97 (47.3%)	58 (47.2%)	0.656
- <i>Requiring hospitalization</i>	590 (52%)	222 (49.7%)	109 (52.7%)	65 (52.8%)	

Abbreviations: TUG = Timed Up and Go; SD = standard deviation; KPS = Karnofsky performance status; bADL = basic activities of daily living; iADL = instrumental activities of daily living

**Table 2.**

The survival rates of patients based on which predicted and actual Timed Up and Go test (TUG).

	<b>60 day survival</b>	<b>180 day survival</b>	<b>360 day survival</b>
Actual TUG <10 seconds	92.9%	88.7%	79.8%
Algorithmic TUG <10 seconds	93.3%	88.8%	80.4%
Actual TUG > 10 seconds	93.2%	85.7%	72.8%
Algorithmic TUG >10 seconds	91.7%	86.0%	73.7%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript