

RESEARCH ARTICLE

Fall Risk Assessment Tools for Elderly Living in the Community: Can We Do Better?

Pierpaolo Palumbo^{1*}, Luca Palmerini¹, Stefania Bandinelli², Lorenzo Chiari¹

1 Department of Electrical, Electronic, and Information Engineering “Guglielmo Marconi”—DEI, University of Bologna, Bologna, Italy, **2** Geriatric Unit, Azienda Sanitaria di Firenze, Florence, Italy

* pierpaolo.palumbo@unibo.it



Abstract

Background

Falls are a common, serious threat to the health and self-confidence of the elderly. Assessment of fall risk is an important aspect of effective fall prevention programs.

Objectives and methods

In order to test whether it is possible to outperform current prognostic tools for falls, we analyzed 1010 variables pertaining to mobility collected from 976 elderly subjects (InCHIANTI study). We trained and validated a data-driven model that issues probabilistic predictions about future falls. We benchmarked the model against other fall risk indicators: history of falls, gait speed, Short Physical Performance Battery (*Guralnik et al.* 1994), and the literature-based fall risk assessment tool FRAT-up (*Cattelani et al.* 2015). Parsimony in the number of variables included in a tool is often considered a proxy for ease of administration. We studied how constraints on the number of variables affect predictive accuracy.

Results

The proposed model and FRAT-up both attained the same discriminative ability; the area under the Receiver Operating Characteristic (ROC) curve (AUC) for multiple falls was 0.71. They outperformed the other risk scores, which reported AUCs for multiple falls between 0.64 and 0.65. Thus, it appears that both data-driven and literature-based approaches are better at estimating fall risk than commonly used fall risk indicators. The accuracy–parsimony analysis revealed that tools with a small number of predictors (~1–5) were suboptimal. Increasing the number of variables improved the predictive accuracy, reaching a plateau at ~20–30, which we can consider as the best trade-off between accuracy and parsimony. Obtaining the values of these ~20–30 variables does not compromise usability, since they are usually available in comprehensive geriatric assessments.

OPEN ACCESS

Citation: Palumbo P, Palmerini L, Bandinelli S, Chiari L (2015) Fall Risk Assessment Tools for Elderly Living in the Community: Can We Do Better? PLoS ONE 10(12): e0146247. doi:10.1371/journal.pone.0146247

Editor: Yue Wang, National Institute for Viral Disease Control and Prevention, CDC, China, CHINA

Received: July 15, 2015

Accepted: December 15, 2015

Published: December 30, 2015

Copyright: © 2015 Palumbo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are from the InCHIANTI study whose authors may be contacted at <http://inchantistudy.net/wp/>

Funding: The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement FARSEEING No. 288940. The InCHIANTI Study (Invecchiare in Chianti, aging in the Chianti area) is currently supported by a grant from the National Institute on Aging (NIH, NIA, Bethesda, USA) and is coordinated by the Tuscany Regional Health Agency in a partnership with the Florence Health Care Agency, the local Administrators and the

primary care physicians of Greve in Chianti and Bagno a Ripoli. The Study was initially managed by the National Institute on Research and Care of the Elderly (INRCA, Ancona, Italy) and it was funded by Italian Health Ministry and by a NIH contract. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Falls are particularly common and burdensome among the elderly. About one third of the older population experiences at least one fall each year [1]. Worldwide, it is estimated that falls are responsible for 35 million disability-adjusted life years [2].

The most widely accepted paradigm for fall prevention in community-dwelling older adults consists of three sequential stages: screening for high fall risk, assessment of multiple risk factors for those at high risk, and implementation of a tailored intervention [3]. The initial screening protocol serves to focus time and financial resources on those subjects at increased risk, and to spare low-risk subjects unnecessary inconvenience. The protocol should be short and easy to administer. The subsequent multifactorial assessment is intended to identify the risk factors to be targeted by the intervention.

Medical societies and health authorities have issued guidelines for screening community-dwelling older people for fall risk. The guidelines from the American and British Geriatric Societies (AGS/BGS update 2011) and the English National Institute for Health and Care Excellence (NICE) propose a combination of simple questions about history of falls in the previous twelve months and difficulties in walking or balance, possibly followed by simple functional tests assessing gait and balance (e.g. Timed Up and Go test (TUG), Performance-Oriented Mobility Assessment, and Berg Balance Test) [4]. The US Centers for Disease Control and Prevention (CDC) combine similar questions and functional tests with a questionnaire (their 'Stay Independent' brochure), which also asks about walking aid use, fear of falling, muscle weakness, proprioception at feet, medications, and depression [5,6]. A previous version of the AGS/BGS guidelines was tested on older disabled women and in community-dwelling older adults and found to be suboptimal with respect to other screening tool and of moderate clinical utility [7]. To the best of our knowledge, no published article reports on the predictive accuracy of current versions of these screening algorithms.

Many other screening tools have been proposed in the literature [3,8–15]. Few of them have been tested outside the derivation cohort. Among those, the TUG has been judged inadequate in several studies [16–18]. Gait speed is an indicator of health state in geriatric populations [19]. Its prognostic value for future falls has been shown to be equivalent to total time to perform the TUG [20]. The Short Physical Performance Battery (SPPB) is a tool to assess physical performance, commonly included in comprehensive geriatric assessments [21,22]. Its association with falls and injurious falls is documented in [23,24]. Its prognostic performance is not reported. History of falls is a strong risk indicator for future falls [3,25,26], although it alone does not suffice for primary prevention. Finally, FRAT-up is a recently proposed predictive tool which issues the probability of falling at least once within the time span of one year [27]. The parameters of FRAT-up were obtained from a systematic review and meta-analysis about risk factors for falls in community-dwelling older people [25]. A comparison between this literature-based approach and data-driven models has never been made.

The first aim of the present study is to test whether a predictive tool, trained using state-of-the-art statistical learning techniques over an extensive dataset, can outperform current tools for fall risk assessment. We trained and tested a statistical model over a dataset pertaining to mobility in a community-dwelling older population in order to obtain an accurate prediction of the number of future falls that a subject will experience. We tested the model on future falls and benchmarked it against these fall risk indicators: history of falls (expressed as number of falls experienced during the twelve months before the assessment), gait speed (usual pace as measured in a 7m walk test), the SPPB summary score, and FRAT-up.

Predictive tools based on a small number of variables could be preferable, as their administration is generally shorter and easier. However, this objective can clash with the need to have

good accuracy in prediction. Both requirements, parsimony and accuracy, should be taken into account in designing a feasible screening tool. The second aim of this study is hence to evaluate the trade-off between the number of variables used within the predictive tool and accuracy of their prediction.

Methods

Data

The dataset comes from the InCHIANTI study ([ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT01331512) NCT01331512), an ongoing population-based cohort study about mobility in the elderly. It consists to date of four waves, about three years apart, initiated in 1999. At each wave the subjects are assessed on a number of different domains and asked about falls experienced in the previous twelve months. The study protocol was approved by the ethical committee of the Italian National Institute of Research and Care of Aging and complies with the Declaration of Helsinki. All participants (or their proxies) received a detailed description of the study purpose and procedures, and gave their written informed consent. More details about the study design and its rationale can be found in [28,29].

We define a sample of the dataset as an assessment of a single person at a specific wave t . The corresponding outcome of the sample to predict is the number of falls as reported by the person at wave $t + 1$. Since each person has participated in up to four waves, there will be up to four samples (and up to three outcomes) for each person. We excluded samples from subjects younger than 65 at the time of the assessment, and samples without information about future falls (e.g. samples from subjects' final assessments or from subjects who did not respond at wave $t + 1$). Thus, we obtained 2313 samples from 976 subjects.

Each subject was assessed on 3280 variables. In addition, from these variables we derived 18 others considered of interest in fall risk assessment (e.g. 'living alone' was derived from questions about social network, 'pain' was obtained from questions about pain in specific body parts). Each variable was manually annotated as either continuous or categorical. We did not consider variables not relevant to the outcome (e.g. date of the home interview, nutritional habits), categorical variables with more than two levels, variables missing more than 50% of their values, and variables where the missing-value-imputation procedure (see section [Model development](#)) did not converge. Thus, the final dataset consisted of 1010 variables. [Table 1](#) provides an overview of the dataset content before and after the procedure for variable selection.

Model development

A probabilistic forecast expresses predictions as a probability distribution over the quantity of interest [30]. We fitted a model that expresses its predictions as negative binomial distributions over the number of falls that a given subject will experience in the following twelve months. The negative binomial distribution is fully specified by two parameters: mean, μ , and dispersion coefficient, θ . Its variance is $\mu + \mu^2 / \theta$. As θ increases, the variance approaches the mean and the negative binomial distribution approaches a Poisson distribution. Accordingly, low values of θ parametrize highly dispersed distributions.

We calculated the mean μ with the output of a Poisson Lasso (least absolute shrinkage and selection operator) regression model [31,32]. The dispersion coefficient θ was calculated from the observed number of falls Y and the predictions μ issued by the regression model, using maximum likelihood and assuming the number of falls as drawn from a negative binomial distribution with mean equal to μ (R function *theta.ml* from package *MASS*[33]). Fitting and evaluation were performed with 10-fold cross-validation. Samples were split into 10 folds in such a way that all the samples from the same subject were assigned to the same fold. In turn, nine

Table 1. Number of variables organized by category before and after the procedure for variable selection.

Area	Brief description	Number of variables	
		Before selection	After selection
Home interview	Mini Mental State Examination (MMSE), activities of daily living (ADL), instrumental activities of daily living (IADL), social network, Centre for Epidemiological Studies Depression Scale (CESD), sleep habits, pain, incontinence, physical activity, falls, fear of falling, shoes, smoking habits	987	167
Clinical visit and disease adjudication	Family and personal clinical history, diagnosed medical conditions	1032	157
Physical exam	FICSIT (Frailty and Injuries Cooperative Studies of Intervention Techniques) balance scale, Purdue Pegboard, Stairs, Repeated Chair Stands, several walking tests, joint range of motion, lower limb muscle power, muscle strength, SPPB summary performance score	602	385
Instrumental exams	ECG, ENG, anthropometric measures, Eco-Color-Doppler, blood pressure, peripheral quantitative computed tomography, bioelectrical impedance analysis	210	111
Laboratory exams	Blood and urine assays	356	103
Medications	Drug classes	93	87

doi:10.1371/journal.pone.0146247.t001

olds were used to fit a Poisson Lasso regression model [32] and calculate a dispersion coefficient. This regression model and the dispersion coefficient were used to issue the probabilistic predictions on the samples of the test fold (Fig 1). In the following, we refer to this approach as the “unconstrained Lasso model”, or “Lasso model” for brevity.

Each fold used a different imputation model to assign missing data [34]; age, sex, and time to walk 7m at self-selected pace served as predictors of missing values. These three variables were chosen because they are associated with health status.

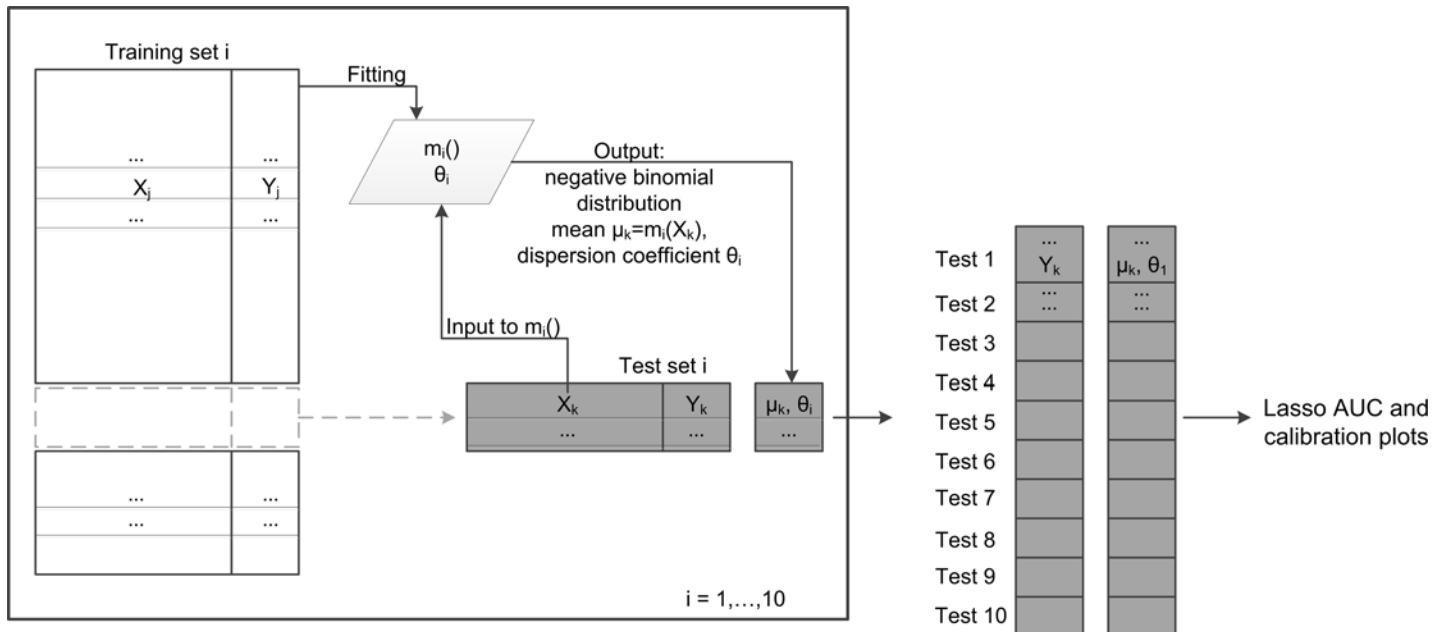
Model benchmark

We compared the performance of the trained model against other well-known fall risk indicators: history of falls [3,25,26], gait speed [20], SPPB summary score [21–24], and FRAT-up risk score [27]. History of falls is the number of falls experienced during the twelve months before the assessment, which in the available dataset ranged from zero to a maximum of nine. Gait speed is measured during a 7m walk test, during which the subjects walk at normal pace. The SPPB summary score is the sum of three sub-scores evaluating gait, balance, and sit-to-stand exercises [21,22]. It ranges from zero to twelve, where higher values indicate better performance. FRAT-up is a web-based tool which computes a risk score which expresses the probability, from zero to one, of experiencing at least one fall within twelve months [27]. History of falls, gait speed, and SPPB were already present in the dataset while the FRAT-up score was calculated as in [27].

Model assessment

We labeled samples (i.e., subjects in a specific wave) as *fallers* if they reported at least one fall at the follow-up after the baseline assessment. Similarly, if they reported more than one fall, they were labeled as *multiple fallers*. All the samples and their associated predictions (history of falls, gait speed, SPPB, FRAT-up, Lasso) were used to evaluate the discriminative ability of the different tools (Fig 1).

We calculated Receiver Operating Characteristic (ROC) curves of the risk scores for fallers and multiple fallers. The ROC curves for the model fitted with Lasso were derived using the means μ of the predictive distributions. The discriminative ability was measured as the area under the ROC curve (AUC). The AUC 95% confidence intervals were calculated via the



X_j : values of variables of sample j
 Y_j : number of falls of sample j
 $m_i(.)$: Poisson regression model fitted with training set i
 θ_i : dispersion coefficient fitted with training set i

Fig 1. 10-fold cross validation scheme. Illustrative scheme of the 10-fold cross validation procedure. During the 10 iterations each sample receives a probabilistic prediction. All these predictions are then used to assess the Lasso model.

doi:10.1371/journal.pone.0146247.g001

DeLong method [35]. The AUCs were compared with Delong tests for paired ROC curves [35].

The Lasso model was also evaluated for calibration (i.e. the agreement between its predictions and the observed number of falls) by means of a reliability diagram, marginal calibration plot, and probability integral transform (PIT). Reliability diagrams (also known as calibration plots or attribute diagrams) are generally used for dichotomous outcomes [36]. Here the reliability diagram was adapted for count data and used to plot the observed fall rate against the predicted fall rate. The marginal calibration plot shows the observed and predicted number of samples for each possible outcome [37]. PIT is used as diagnostics of probabilistic calibration. It detects whether the variance of the probabilistic predictions agrees with the dispersion of the observations (neutral dispersion), or whether it expresses too little or too much uncertainty (under-dispersion or over-dispersion, respectively) [30]. It was calculated according to the non-randomized procedure for count data described in [37].

Accuracy-parsimony analysis

The Lasso regression performs variable selection and parameter estimation at the same time. It encourages sparse solutions, i.e. solutions that make use of a small number of variables [31,38]. In order to study how the parsimony of the model affects its predictive accuracy, we studied the performance of the model when fitted under a constraint on the maximum number of variables, n , to be included. We refer to this approach as “constrained Lasso model”.

In particular, for a given n , we trained the constrained Lasso model according to the 10-fold cross-validation procedure described above. We calculated the mean number of variables actually included in the models averaging over the 10 regression models (one for each fold of the

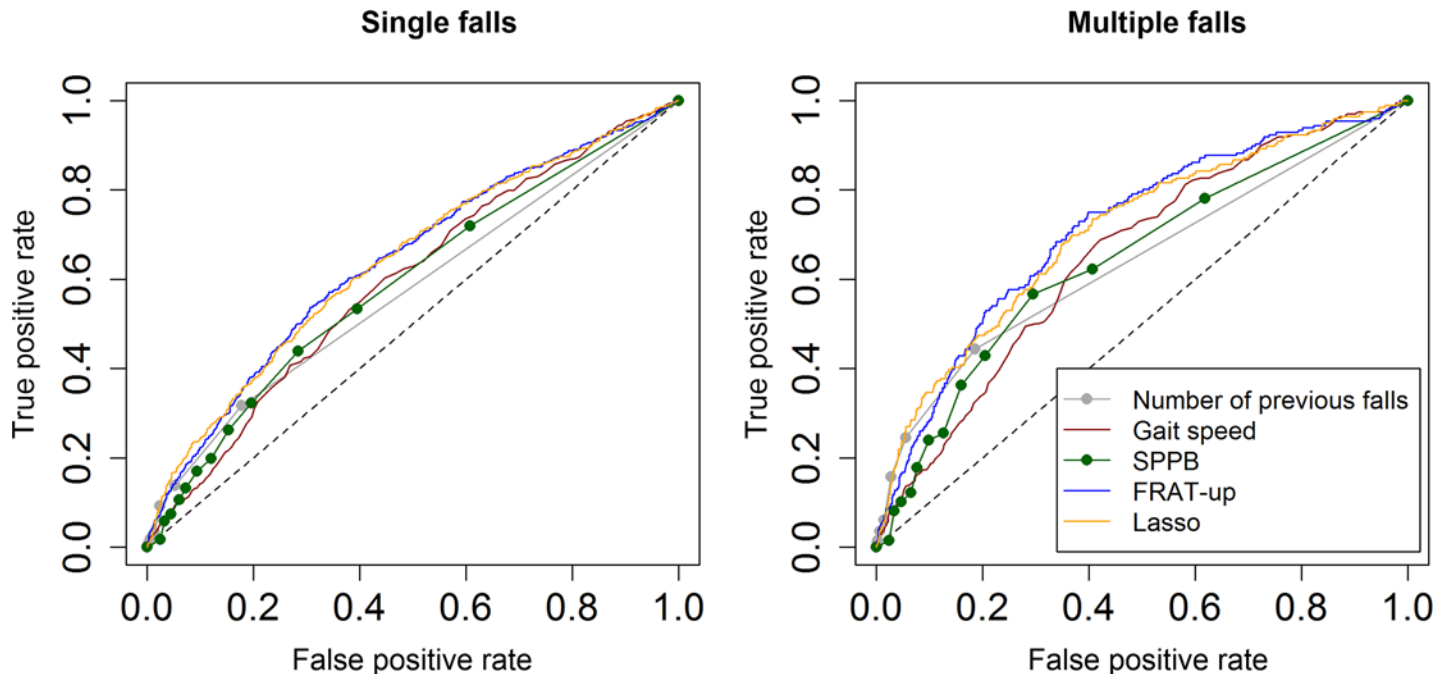


Fig 2. ROC curves. ROC curves of the Lasso model and the other risk indicators (number of previous falls, gait speed, SPPB, and FRAT-up) when predicting for single (left panel) and multiple (right panel) falls.

doi:10.1371/journal.pone.0146247.g002

cross-validation) fitted during the cross-validation procedure. We evaluated the accuracy of the model with AUCs for single and multiple fallers, and mean squared error (MSE). MSE was calculated as the mean squared difference between the observed number of falls and the predicted fall rate μ . We repeated this analysis varying n from 1 to 40.

Results

The ten unconstrained Lasso models fitted within the procedure of 10-fold cross-validation account for a number of variables that ranges from 21 to 41, with a mean of 29.4. Details about which variables were selected more frequently, and their regression coefficients, are given in [S1 Table](#).

ROC curves of the five risk scores for single and multiple fallers are shown in [Fig 2](#). The associated AUCs and the results of the hypothesis tests for paired ROC curves are reported in [Table 2](#).

[Fig 3](#) shows an example of the output of the Lasso model for four representative samples at the 2.5th, 10th, 90th, and 97.5th percentiles of the Lasso risk score, compared with the distribution of the observed number of falls in the InCHIANTI dataset. The expected number of falls for the four selected samples is 0.21, 0.23, 0.66, 1.08, respectively. The fall rate in the InCHIANTI population (baseline data) is 0.42 falls/(person · year).

The Lasso model's reliability diagram, marginal calibration plot, and histogram of the PIT are given in [Fig 4](#). Results for the assessment of marginal calibration are shown in more detail in [S2 Table](#).

[Fig 5](#) reports the results of the accuracy-parsimony analysis. The mean number of variables for the ten constrained regression models is less than the maximum number of variables n set initially by the constraint. As we relax the constraint (i.e. as n increases), the mean number of variables included in the models increases, and the predictive accuracy (measured with MSE

Table 2. Discriminative ability of five fall risk indicators. Comparisons with FRAT-up and Lasso model were made with DeLong tests for paired AUCs.

	Single falls		Multiple falls	
	AUC (95% C.I.)	p value. Risk indicator vs FRAT-up / Risk indicator vs Lasso	AUC (95% C.I.)	p value. Risk indicator vs FRAT-up / Risk indicator vs Lasso
Number of previous falls	0.574 (0.551–0.597)	** / **	0.640 (0.603–0.678)	** / **
Gait speed	0.594 (0.566–0.622)	** / **	0.653 (0.615–0.692)	** / *
SPPB	0.590 (0.563–0.618)	** / **	0.645 (0.604–0.686)	** / **
FRAT-up	0.638 (0.610–0.666)	– / 0.92	0.713 (0.675–0.752)	– / 0.62
Lasso	0.639 (0.611–0.667)	0.92 / –	0.708 (0.669–0.747)	0.62 / –

* = p<0.01
 ** = p<0.001.

doi:10.1371/journal.pone.0146247.t002

and AUC for single and multiple fallers) improves until a plateau is reached. For the AUC this occurs at about 20 variables; for the MSE at about 30 variables.

Discussion

In this study we have developed a model for fall prediction using a dataset that is large, in terms of number of samples and number of variables related to mobility. We have assessed its predictive properties and benchmarked it against four other risk indicators. We have further investigated whether and to what degree the parsimony of the model compromises its predictive accuracy.

The results show that the AUCs of the Lasso model and FRAT-up are similar and significantly higher than the other risk scores. FRAT-up parameters were derived from the literature [25,27], while the approach proposed here is strongly data-driven. The equivalence of discriminative ability between Lasso and FRAT-up confirms the validity of the literature-based approach.

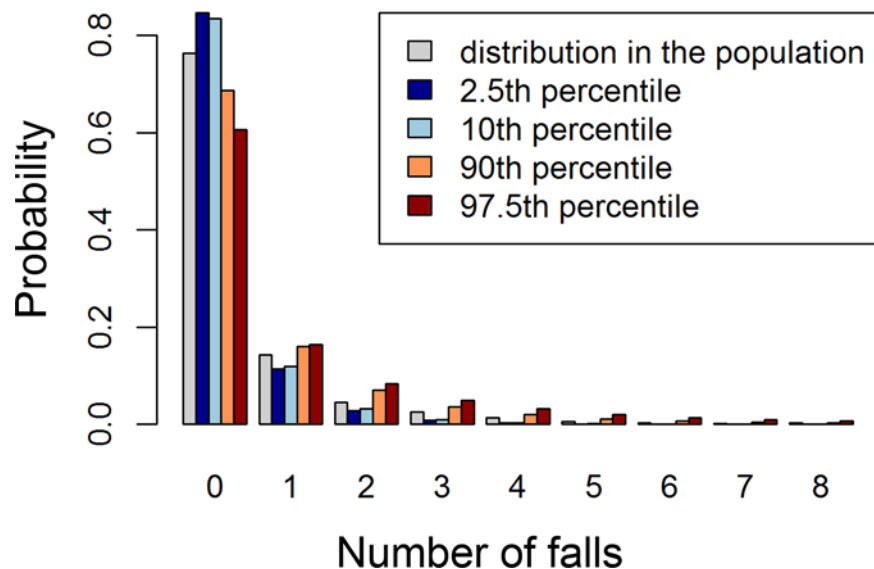


Fig 3. Representative distributions of the Lasso model. Histogram showing the predictive distributions (probability mass functions) on the number of falls for four samples at the 2.5th, 10th, 90th, and 97.5th percentiles of the Lasso risk score. These are compared with the distribution of the number of falls in the InCHIANTI population (in gray).

doi:10.1371/journal.pone.0146247.g003

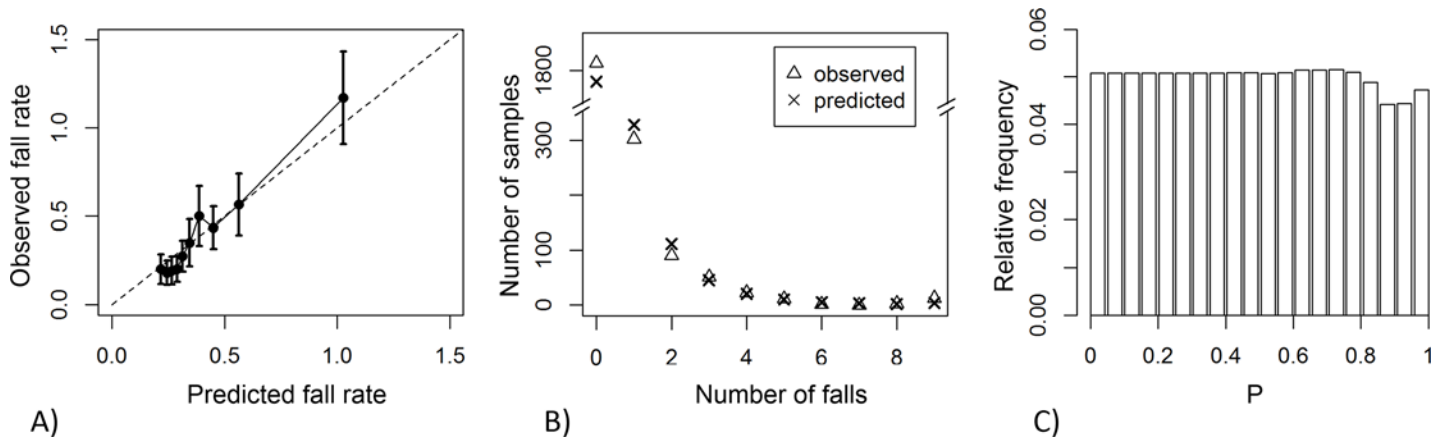


Fig 4. Assessment of Lasso model calibration. (A) Reliability diagram: observed vs predicted fall rate, obtained from grouping samples according to deciles on the risk score; error bars indicate 95% confidence intervals. (B) Marginal calibration plot: observed and predicted number of samples vs number of falls. (C) Histogram of the probability integral transform.

doi:10.1371/journal.pone.0146247.g004

More analyses presented in [S1 File](#) show that the possible hazards of training statistical models in high-dimensional spaces have been avoided. In particular, the learning curves show that more samples for training will not lead to substantial improvements in AUC or MSE. Higher values of AUC could instead be reached from different sources of information, e.g. from variables derived from wearable inertial sensors [12]. AUCs for multiple falls are higher than AUCs for single falls. This result is consistently found for all the risk indicators, and was already reported by other empirical [39] and modeling studies [40].

The Lasso model is well calibrated (Fig 4). Calibration refers to different properties of statistical consistency between predictions and observations [30,41]. The reliability diagram (Fig 4A) shows that the number of predicted falls agrees with the number of observed falls on

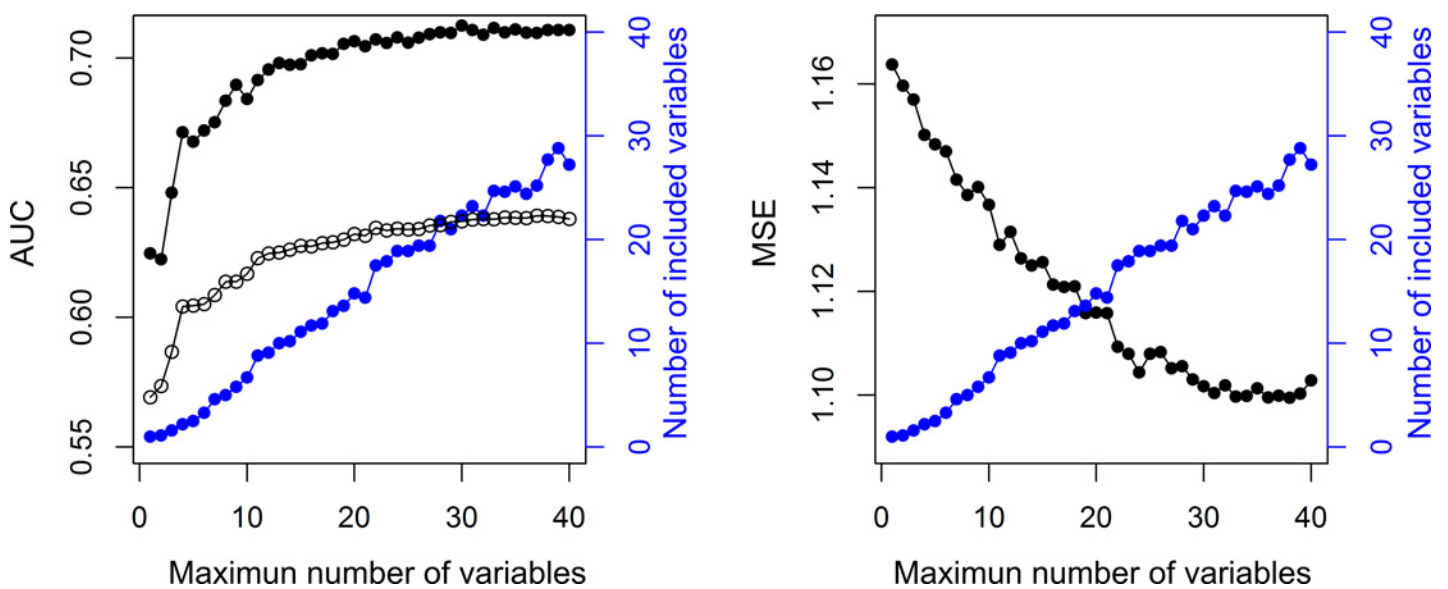


Fig 5. Accuracy-parsimony analysis. Performance of the model when constraining the maximum number of variables to be included in the model. Left panel, left axis: AUC for single falls (black empty circles), AUC for multiple falls (black filled circles). Right panel, left axis: MSE (black filled circles). Both panels, right axes: mean number of variables that were actually included in the models (blue circles).

doi:10.1371/journal.pone.0146247.g005

samples grouped in deciles of the risk score. It also shows that the fall rate is constant across the first 4 deciles (40% of the samples). This represents poor discrimination among those at low risk, since increasing values of risk do not correspond to increasing values of observed falls. The marginal calibration plot (Fig 4B) shows that the model performs fairly well in predicting how many samples will demonstrate a given number of falls (see also S2 Table). The histogram of the PIT (Fig 4C) shows that the model is neutrally dispersed; that is, the negative binomial distributions (meant to express the predictions) have a variance that reflects the right amount of uncertainty about the number of falls that the subject will experience [37]. Conversely, Poisson predictions, obtained without calculating the dispersion coefficient, substantially underestimated the number of non-fallers and exhibited under-dispersion (S1 Fig). Given the unexplained variance in fall incidence across different studies [42], however, the good calibration properties obtained using the InCHIANTI dataset are not guaranteed to hold on other datasets.

Fall risk indicators such as gait speed, SPPB score, and history of falls can be interpreted as performance scores: the lower (or higher, in the case of history of falls) their values, the higher the risk of falling. In contrast, FRAT-up and the Lasso model provide probabilistic predictions. FRAT-up outputs the probability of falling at least once during the twelve months after the assessment; the Lasso model supplies the probability distribution of the number of falls that will be experienced during the same time span. Predicting the number of falls instead of a dichotomous outcome (whether at least one or two falls will occur) provides more information without drawbacks.

Expressing a prediction in probabilistic terms has advantages compared to simpler risk indicators. First, it is possible to aggregate and compare probabilistic predictions issued by different tools for the same health outcome; moreover, the risks of different health outcomes can be compared. Second, calibrated probabilistic models provide accurate statements about groups of subjects. Third, since the prediction is expressed as a probability distribution, its interpretation does not rely on any specific knowledge. In contrast, interpreting other scores requires tool-specific knowledge (e.g. their admissible range and whether they are positively or negatively oriented). Multidisciplinary research is currently investigating how to best express predictions in order to facilitate doctor-patient communication, convey the uncertainty associated with the prediction, and the most influential determinants [43–46]. In this regard, a figure that compares the patient's issued prediction with reference distributions from the general population (see, e.g., Fig 3) could be a viable option to clearly convey the result of the risk assessment together with its associated uncertainty.

The variables that were selected most frequently in the 10-fold validation procedure (S1 Table) are primarily known risk factors and indicators for fall risk (e.g. history of falls, self-perceived health status, depression, number of medications, and use of drugs acting on the central nervous system) [25]. With reference to the categorization of fall risk factors proposed in [1], most of the selected variables are biological risk factors, and few are behavioral or socio-economic. None is environmental, since the dataset does not contain information on this kind of risk factors. Other selected variables (e.g. use of the antibacterial quinolone) are unexpected. Indeed, it is known that Lasso accidentally selects 'noisy' variables. Although other regression techniques alleviate this problem, they are more computationally demanding and have failed to prove better predictive accuracy in situations of low signal-to-noise ratio [47].

The accuracy-parsimony analysis shows that predictive accuracy improves as the number of variables increases, up to 20–30. This result, in line with the multifactorial etiology of falls, may explain why screening tools employing a very small number of variables have suboptimal performance. The AGS/BGS guidelines suggest a cascade of three to six questions and simple assessments: two or more falls in the previous twelve months, acute fall, self-reported

difficulties, and assessed abnormalities in gait and balance. The CDC's 'Stay independent' brochure [5] contains twelve questions. Its score is integrated with three questions asked directly by the clinician, and possibly with an assessment of gait, balance, and strength [6]. The FRAT-up questionnaire is made of 28 items, with the possibility of leaving some fields blank because it embeds prevalence information on individual risk factors [48]. Tools that require a high number of input variables can be long and expensive to administer if all the information has to be collected de novo. However, much of the information about fall risk factors is already available to the physician, from the geriatric comprehensive assessment [49]. As a result, a fall risk evaluation integrated into the geriatric comprehensive assessment could provide a good prediction without imposing an additional burden on subjects under assessment or their healthcare professionals.

The InCHIANTI dataset allowed us to make an extensive search on different domains related to mobility and falls in the elderly. However, we have to acknowledge some main limitations of the present study. First, we did not have information about fall hazards in the environment. Second, we excluded all the categorical variables with more than two levels from our analysis. Our choice was driven by the will to avoid a further increase in the number of variables, since fitting a regression model means each categorical variable with p levels is replaced with $p - 1$ dummy variables. Indeed, this exclusion criterion may have led to the loss of potentially valuable predictors (e.g. variables for the diagnosis for some diseases, which allow values such as 'Yes', 'No', and 'Diagnosis not definite'). Lastly, it must be made clear that in this paper predictions have been validated on falls occurring during a period of one year starting two years after the risk factor assessment. This was due to the study design, as described in section Data. This may be one of the causes of the relatively low values achieved on the AUC.

Conclusions

We have presented the development and assessment of a tool that issues probabilistic predictions on the number of future falls in a cohort of community-dwelling older subjects. We have trained this model over a dataset that is large, in terms of both number of variables related to mobility and falls and number of samples. We have benchmarked it against other risk indicators. The trained model and FRAT-up outperformed simple fall risk indicators. Despite the breadth of the dataset and the use of state-of-the-art tools of statistical learning, the trained model was not able to reach a better discriminative ability than FRAT-up. This finding supports the validity of the literature-based approach used to develop FRAT-up. Both the data-driven and literature-based approaches are better at estimating fall risk than commonly used fall risk indicators.

The accuracy-parsimony analysis has shown that predictive accuracy improves as the number of variables increases up to 20–30. This suggests that fall prediction is more accurate when based on multiple fall risk factors and indicators; thus simplistic screening tests (three to six variables) are suboptimal in terms of predictive accuracy. Since common risk factors and indicators are already part of geriatric comprehensive assessments, integrating prognostic tools for falls into them could improve the prediction without compromising usability.

Supporting Information

S1 Fig. Probability integral transform for Poisson and negative binomial predictions. Histograms of the probability integral transforms for Poisson (left panel) and negative binomial (right panel) predictions. The U shape for the predictions expressed in terms of Poisson distributions indicates under-dispersion, i.e. understatement of the uncertainty associated with the prediction. The flat shape for the predictions expressed as negative binomial distributions

indicates neutral-dispersion.
(TIF)

S1 File. Learning curves. Performance of the Lasso prediction as a function of the number of samples used for training.
(DOCX)

S1 Table. Variables selected in the Lasso regression. Variables that were selected more frequently in the 10-fold validation procedure and their standardized regression coefficients.
(DOCX)

S2 Table. Marginal calibration assessment. Observed and predicted number of samples reporting a given number of falls. $Error = predicted - observed$. The marginal calibration plot is presented in [Fig 4B](#).
(DOCX)

Acknowledgments

The authors thank Angelica Maria Perez Oviedo and Michael Cuba Sotelo for help in manually annotating the dataset, Luca Cattelani and Federico Chesani for fruitful discussions, Marco Colpo for reading and commenting the manuscript, Kristina Mayberry for help with language revision.

Author Contributions

Conceived and designed the experiments: PP LP LC SB. Performed the experiments: PP LP LC SB. Analyzed the data: PP LP. Contributed reagents/materials/analysis tools: SB. Wrote the paper: PP LP LC.

References

1. World Health Organization. WHO global report on falls prevention in older age. Geneva, Switzerland: World Health Organization; 2007.
2. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012; 380: 2197–223. doi: [10.1016/S0140-6736\(12\)61689-4](https://doi.org/10.1016/S0140-6736(12)61689-4) PMID: [23245608](https://pubmed.ncbi.nlm.nih.gov/23245608/)
3. Ganz DA, Bao Y, Shekelle PG, Rubenstein LZ. Will my patient fall? *JAMA*. 2007; 297: 77–86. doi: [10.1001/jama.297.1.77](https://doi.org/10.1001/jama.297.1.77) PMID: [17200478](https://pubmed.ncbi.nlm.nih.gov/17200478/)
4. Panel of Falls in Older Persons American Geriatrics Society and British Geriatrics Society. Summary of the Updated American Geriatrics Society/British Geriatrics Society clinical practice guideline for prevention of falls in older persons. *J Am Geriatr Soc*. 2011; 59: 148–57. doi: [10.1111/j.1532-5415.2010.03234.x](https://doi.org/10.1111/j.1532-5415.2010.03234.x) PMID: [21226685](https://pubmed.ncbi.nlm.nih.gov/21226685/)
5. Centers for Disease Control and Prevention. Stay independent brochure [Internet]. [cited 21 Feb 2015]. Available: http://www.cdc.gov/homeandrecreationalafety/pdf/steady/stay_independent.pdf
6. Centers for Disease Control and Prevention. Algorithm for fall risk assessment & intervention [Internet]. [cited 21 Feb 2015]. Available: http://www.cdc.gov/homeandrecreationalafety/pdf/steady/algorithm_fall_risk_assessment.pdf
7. Muir SW, Berg K, Chesworth B, Klar N, Speechley M. Application of a fall screening algorithm stratified fall risk but missed preventive opportunities in community-dwelling older adults: A prospective study. *J Geriatr Phys Ther*. 2010; 33: 165–172. doi: [10.1097/JPT.0b013e3181ff23cc](https://doi.org/10.1097/JPT.0b013e3181ff23cc) PMID: [21717920](https://pubmed.ncbi.nlm.nih.gov/21717920/)
8. Fabre JM, Ellis R, Kosma M, Wood RH. Literature Review Falls Risk Factors and a Compendium of Falls Risk Screening Instruments. *J Geriatr Phys Ther*. 2010; 33: 184–197. doi: [10.1097/JPT.0b013e3181ff2a24](https://doi.org/10.1097/JPT.0b013e3181ff2a24) PMID: [21717922](https://pubmed.ncbi.nlm.nih.gov/21717922/)
9. 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 Rehabil Res Dev*. 2008; 45: 1105–16. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19235113> PMID: [19235113](https://pubmed.ncbi.nlm.nih.gov/19235113/)

10. Close JCT, Lord SR. Fall assessment in older people. *BMJ*. 2011;343. doi: [10.1136/bmj.d5153](https://doi.org/10.1136/bmj.d5153)
11. Perell KL, Nelson A, Goldman RL, Luther SL, Prieto-Lewis N, Rubenstein LZ. Fall risk assessment measures: an analytic review. *J Gerontol A Biol Sci Med Sci*. 2001; 56: M761–6. Available: <http://www.ncbi.nlm.nih.gov/pubmed/11723150> PMID: [11723150](https://pubmed.ncbi.nlm.nih.gov/11723150/)
12. Howcroft J, Kofman J, Lemaire ED. Review of fall risk assessment in geriatric populations using inertial sensors. *J Neuroeng Rehabil. Journal of NeuroEngineering and Rehabilitation*; 2013; 10: 91. doi: [10.1186/1743-0003-10-91](https://doi.org/10.1186/1743-0003-10-91)
13. Persad CC, Cook S, Giordani B. Assessing falls in the elderly: should we use simple screening tests or a comprehensive fall risk evaluation? *Eur J Phys Rehabil Med*. 2010; 46: 249–59. Available: <http://www.ncbi.nlm.nih.gov/pubmed/20485227> PMID: [20485227](https://pubmed.ncbi.nlm.nih.gov/20485227/)
14. Scott V, Votova K, Scanlan A, Close J. Multifactorial and functional mobility assessment tools for fall risk among older adults in community, home-support, long-term and acute care settings. *Age Ageing*. 2007; 36: 130–139. doi: [10.1093/ageing/af1165](https://doi.org/10.1093/ageing/af1165) PMID: [17293604](https://pubmed.ncbi.nlm.nih.gov/17293604/)
15. Lee J, Geller AI, Strasser DC. Analytical review: focus on fall screening assessments. *PM R*. 2013; 5: 609–21. doi: [10.1016/j.pmrj.2013.04.001](https://doi.org/10.1016/j.pmrj.2013.04.001) PMID: [23880047](https://pubmed.ncbi.nlm.nih.gov/23880047/)
16. Barry E, Galvin R, Keogh C, Horgan F, Fahey T. Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta- analysis. *BMC Geriatr*. 2014; 14: 14. doi: [10.1186/1471-2318-14-14](https://doi.org/10.1186/1471-2318-14-14) PMID: [24484314](https://pubmed.ncbi.nlm.nih.gov/24484314/)
17. Schoene D, Wu SM-S, Mikolaizak AS, Menant JC, Smith ST, Delbaere K, et al. Discriminative ability and predictive validity of the timed up and go test in identifying older people who fall: systematic review and meta-analysis. *J Am Geriatr Soc*. 2013; 61: 202–8. doi: [10.1111/jgs.12106](https://doi.org/10.1111/jgs.12106) PMID: [23350947](https://pubmed.ncbi.nlm.nih.gov/23350947/)
18. Beauchet O, Fantino B, Allali G, Muir SW, Montero-Odasso M, Annweiler C. Timed up and go test and risk of falls in older adults: A systematic review. *J Nutr Health Aging*. 2011; 15: 933–938. doi: [10.1007/s12603-011-0062-0](https://doi.org/10.1007/s12603-011-0062-0) PMID: [22159785](https://pubmed.ncbi.nlm.nih.gov/22159785/)
19. Abellan van Kan G, Rolland Y, Andrieu S, Bauer J, Beauchet O, Bonnefoy M, et al. Gait speed at usual pace as a predictor of adverse outcomes in community-dwelling older people an International Academy on Nutrition and Aging (IANA) Task Force. *J Nutr Health Aging*. 2009; 13: 881–9. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19924348> PMID: [19924348](https://pubmed.ncbi.nlm.nih.gov/19924348/)
20. Viccaro LJ, Perera S, Studenski S a. Is timed up and go better than gait speed in predicting health, function, and falls in older adults? *J Am Geriatr Soc*. 2011; 59: 887–92. doi: [10.1111/j.1532-5415.2011.03336.x](https://doi.org/10.1111/j.1532-5415.2011.03336.x) PMID: [21410448](https://pubmed.ncbi.nlm.nih.gov/21410448/)
21. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol*. 1994;
22. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. *N Engl J Med*. 1995; 332: 556–61. doi: [10.1056/NEJM199503023320902](https://doi.org/10.1056/NEJM199503023320902) PMID: [7838189](https://pubmed.ncbi.nlm.nih.gov/7838189/)
23. Veronese N, Bolzetta F, Toffanello ED, Zambon S, De Rui M, Perissinotto E, et al. Association between Short Physical Performance Battery and falls in older people: the Progetto Veneto Anziani Study. *Rejuvenation Res*. 2014; 17: 276–84. doi: [10.1089/rej.2013.1491](https://doi.org/10.1089/rej.2013.1491) PMID: [24387140](https://pubmed.ncbi.nlm.nih.gov/24387140/)
24. Ward RE, Leveille SG, Beauchamp MK, Trivison T, Alexander N, Jette AM, et al. Functional performance as a predictor of injurious falls in older adults. *J Am Geriatr Soc*. 2015; 63: 315–20. doi: [10.1111/jgs.13203](https://doi.org/10.1111/jgs.13203) PMID: [25688606](https://pubmed.ncbi.nlm.nih.gov/25688606/)
25. Deandrea S, Lucenteforte E, Bravi F, Foschi R, La Vecchia C, Negri E. Risk Factors for Falls in Community-dwelling Older People: A Systematic Review and Meta-analysis. *Epidemiology*. 2010; 21: 658–668. doi: [10.1097/EDE.0b013e3181e89905](https://doi.org/10.1097/EDE.0b013e3181e89905) PMID: [20585256](https://pubmed.ncbi.nlm.nih.gov/20585256/)
26. Ambrose AF, Paul G, Hausdorff JM. Risk factors for falls among older adults : A review of the literature. *Maturitas*. Elsevier Ireland Ltd; 2013; 75: 51–61. doi: [10.1016/j.maturitas.2013.02.009](https://doi.org/10.1016/j.maturitas.2013.02.009) PMID: [23523272](https://pubmed.ncbi.nlm.nih.gov/23523272/)
27. Cattalani L, Palumbo P, Palmerini L, Bandinelli S, Becker C, Chesani F, et al. FRAT-up, a web-based fall risk assessment tool for elderly people living in the community. *J Med Internet Res*. 2015; 17: e41. doi: [10.2196/jmir.4064](https://doi.org/10.2196/jmir.4064) PMID: [25693419](https://pubmed.ncbi.nlm.nih.gov/25693419/)
28. Ferrucci L, Bandinelli S, Benvenuti E, Di Iorio A, Macchi C, Harris TB, et al. Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *J Am Geriatr Soc*. 2000; 48: 1618–25. Available: <http://www.ncbi.nlm.nih.gov/pubmed/11129752> PMID: [11129752](https://pubmed.ncbi.nlm.nih.gov/11129752/)
29. InCHIANTI. InCHIANTI Study door [Internet]. [cited 6 Jul 2013]. Available: <http://inchantistudy.net/>
30. Gneiting T, Katzfuss M. Probabilistic Forecasting. *Annu Rev Stat Its Appl*. 2014; doi: [10.1146/annurev-statistics-062713-085831](https://doi.org/10.1146/annurev-statistics-062713-085831)

31. Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B (Statistical Methodol)*. 1996; 58: 267–288.
32. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010; 33: 1–22. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2929880&tool=pmcentrez&rendertype=abstract> PMID: 20808728
33. Venables VN, Ripley BD. *Modern Applied Statistics with S* [Internet]. Fourth. New York: Springer; 2002. Available: <http://www.stats.ox.ac.uk/pub/MASS4>
34. van Buuren S, Groothuis-oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw*. 2011; 45.
35. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011; 12: 77. doi: [10.1186/1471-2105-12-77](https://doi.org/10.1186/1471-2105-12-77) PMID: 21414208
36. Murphy AH, Winkler RL. Reliability of subjective probability forecasts of precipitation and temperature. In: *Journal of the Royal Society. Series C (Applied Statistics)* [Internet]. 1977 [cited 11 Feb 2015] pp. 41–47. Available: http://www.jstor.org/stable/2346866?seq=2#page_scan_tab_contents
37. Czado C, Gneiting T, Held L. Predictive model assessment for count data. *Biometrics*. 2009; 65: 1254–61. doi: [10.1111/j.1541-0420.2009.01191.x](https://doi.org/10.1111/j.1541-0420.2009.01191.x) PMID: 19432783
38. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: Data Mining, Inference, and Prediction*. Second ed. Springer; 2009.
39. Nevitt MC, Cummings SR, Kidd S, Black D. Risk factors for recurrent nonsyncopal falls. A prospective study. *JAMA*. 1989; 261: 2663–8. Available: <http://www.ncbi.nlm.nih.gov/pubmed/2709546> PMID: 2709546
40. Palumbo P, Palmerini L, Chiari L. A probabilistic model to investigate the properties of prognostic tools for falls. *Methods Inf Med*. 2015; 54.
41. Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. *J R Stat Soc Ser B*. 2007; 69: 243–268.
42. Rapp K, Freiberger E, Todd C, Klenk J, Becker C, Denking M, et al. Fall incidence in Germany: results of two population-based studies, and comparison of retrospective and prospective falls data collection methods. *BMC Geriatr*. 2014; 14: 105. doi: [10.1186/1471-2318-14-105](https://doi.org/10.1186/1471-2318-14-105) PMID: 25241278
43. Joslyn S, Leclerc J. Decisions with uncertainty: the glass half full. *Curr Dir Psychol Sci*. 2013; 22: 308–315. doi: [10.1177/0963721413481473](https://doi.org/10.1177/0963721413481473)
44. Spiegelhalter D, Pearson M, Short I. Visualizing uncertainty about the future. *Science*. 2011; 333: 1393–400. doi: [10.1126/science.1191181](https://doi.org/10.1126/science.1191181) PMID: 21903802
45. Edwards A. Explaining risks: turning numerical data into meaningful pictures. *BMJ*. 2002; 324: 827–830. doi: [10.1136/bmj.324.7341.827](https://doi.org/10.1136/bmj.324.7341.827) PMID: 11934777
46. Smith AF. Discussion of risk pervades doctor-patient communication. *BMJ*. 2002; 325: 548–548. doi: [10.1136/bmj.325.7363.548](https://doi.org/10.1136/bmj.325.7363.548)
47. Zou H. The Adaptive Lasso and Its Oracle Properties. *J Am Stat Assoc*. 2006; 101: 1418–1429. doi: [10.1198/016214506000000735](https://doi.org/10.1198/016214506000000735)
48. Cattalani L, Palumbo P, Palmerini L, Chesani F, Chiari L. FRAT-up Web Application [Internet]. 2014. Available: <http://ffrat.farseeingresearch.eu/>
49. Stuck AE, Iliffe S. Comprehensive geriatric assessment for older adults. *BMJ*. 2011; 343: d6799. doi: [10.1136/bmj.d6799](https://doi.org/10.1136/bmj.d6799) PMID: 22034147