A Probabilistic Model to Investigate the Properties of Prognostic Tools for Falls*

P. Palumbo, L. Palmerini, L. Chiari
Department of Electrical, Electronic, and Information Engineering – Guglielmo Marconi (DEI), Bologna, Italy

Keywords
Elderly, fall, poisson, risk assessment, probabilistic model, prognostic tool

Summary
Background: Falls are a prevalent and burdensome problem in the elderly. Tools for the assessment of fall risk are fundamental for fall prevention. Clinical studies for the development and evaluation of prognostic tools for falls show high heterogeneity in the settings and in the reported results. Newly developed tools are susceptible to over-optimism.

Objectives: This study proposes a probabilistic model to address critical issues about fall prediction through the analysis of the properties of an ideal prognostic tool for falls.

Methods: The model assumes that falls occur within a population according to the Greenwood and Yule scheme for accident-proneness. Parameters for the fall rate distribution are estimated from counts of falls of four different epidemiological studies.

Results: We obtained analytic formulas and quantitative estimates for the predictive and discriminative properties of the ideal prognostic tool. The area under the receiver operating characteristic curve (AUC) ranges between about 0.80 and 0.89 when prediction on any fall is made within a follow-up of one year. Predicting on multiple falls results in higher AUC.

Conclusions: The discriminative ability of current validated prognostic tools for falls is sensibly lower than what the proposed ideal perfect tool achieves. A sensitivity analysis of the predictive and discriminative properties of the tool with respect to study settings and fall rate distribution identifies major factors that can account for the high heterogeneity of results observed in the literature.

1. Introduction
Falls in the elderly are a major cause of morbidity and mortality. About 30% of the population aged 65 or more fall at least once each year [1] and the annual fall rate in this age group is about 0.65 falls per person [2]. The fall risk increases with age and is higher in populations of older adults hospitalized or living in long-term care institutions. Physical injuries due to falls account for 40% of all injury deaths [1]. Even when no overt physical impairment is reported after a fall, loss of self-confidence and other psychological consequences can lead to restrictions in physical activity and social participation.

A fundamental element for each preventive strategy is the early identification of persons at risk, thus the availability of prognostic tools for fall risk assessment. Besides assessment tools whose principal aim is to identify key risk factors to target factor-specific interventions [3, 4] – which will not be dealt with in this article – common prognostic tools either produce a continuum score related to the probability of occurrence of one or more falls in a given amount of time [5], or are simple categorical prediction tools that stratify the population according to a dichotomous “low-versus-high risk” logic [6]. Systematic reviews have generally remarked that: • only few tools have been externally validated [7] (i.e. have been evaluated in a population different from the one employed in the development phase, see [8, 9] for a comprehensive discussion about validation of prognostic models); • not a single validated tool has so far shown excellent discriminative properties [6, 7, 10]; • heterogeneity of population characteristics and study settings affect the predictive properties of the tools [6, 11, 12].

From the literature in biostatistics and epidemiology it is known that even a ‘perfect’ prognostic tool, i.e. a tool that assigns each subject their true probability to develop the outcome of interest [13], cannot reach perfect discrimination [14, 15], and that the upper limit for the area under the receiver operating characteristic (ROC) curve (AUC) depends on the distribution of risk in the population [13, 16]. Nevertheless the clinical literature about falls has never discussed its results in light of these theoretical considerations and some recently-proposed tools incorporating inertial sensors data [17] have shown good but unlikely results. Furthermore, it is known that if a clinical or biological marker is used to predict the time until the development of a given outcome, the ROC curve of the associated prognostic tool is dependent on...
the censoring time [18]. Finally, the studies that have developed or evaluated fall risk assessment tools have considered as outcome of interest the condition of having fallen either at least once (e.g. [19]) or at least twice during the follow-up (e.g. [20]). How the follow-up length of a prospective study and the definition of the outcome of interest impact on the estimated predictive properties of the fall risk tool under evaluation has not been investigated yet.

2. Objectives

By means of a probabilistic model we aim to investigate the above mentioned theoretical issues about fall prediction. Setting a framework where an ideal prognostic tool for fall risk is evaluated in a population enrolled in a hypothetical prospective clinical study, we aim to derive analytically and evaluate quantitatively its predictive and discriminative performances, and to investigate how these performances are affected by i) the distribution of the fall rate in the population, ii) the follow-up duration, and iii) the definition of faller as single faller or multiple faller.

3. Methods

3.1 Probabilistic Model

We assume to evaluate a prognostic tool for falls in an infinite population within a prospective study. All the subjects \( \omega_1, \omega_2, \ldots \) of this population are followed over time, from \( t = 0 \), instant of the baseline, until \( t = \tau \), duration of the follow-up.

To each subject \( \omega_i \) of the population, we associate two random variables, namely \( N_i \) and \( \Lambda_i \). \( N_i \) is a random variable accounting for the number of falls that \( \omega_i \) will experience during the follow-up. Conditional on \( \Lambda_i \), \( N_i \) is assumed independent of any other random variable of the model. With this assumption the number of falls that a subject \( \omega_i \) will experience during the follow-up depends only on the value of \( \Lambda_i \), \( \Lambda_i \) is the expected fall rate of \( \omega_i \) (expressed as the number of falls per year). Then, \( \Lambda_i \) is interpreted as a measure of the proneness to falling of \( \omega_i \). We assume the random variables \( \Lambda_1, \Lambda_2, \ldots \) to be independent and identically distributed according to a distribution \( F \).

Thus, \( (N_1, \Lambda_1), (N_2, \Lambda_2), \ldots \) are independent and identically distributed (i.i.d.) couples of random variables. We call \( (N, \Lambda) \) one of these i.i.d. couples. We assume that \( N \) has a conditional Poisson distribution with mean \( \lambda \tau \) given \( \Lambda = \lambda \). Its conditional probability mass function is thus:

\[
g(n; \lambda \tau) = \frac{\lambda^n}{n!} e^{-\lambda \tau} \quad \text{for } n = 0, 1, 2, \ldots
\]

The marginal probability mass function for \( N \) is clearly:

\[
h(n) = \int_0^{\infty} g(n; \lambda \tau) dF(\lambda)
\]

Equation 2 says that \( N \) follows a mixture of Poisson distributions, with mixing distribution \( F \). Thus fall counts are regarded as arising from a mixture of subjects, each falling according to a Poisson law conditioned on their expected fall rate, with the expected fall rate being distributed according to \( F \). As supported by empirical evidence [21], we hypothesize the marginal distribution for \( N \) to be a negative binomial:

\[
N \sim \text{NegBin}(k, \mu)
\]

with \( k \) and \( \mu \) being the two real and positive parameters that characterize the distribution. Its probability mass function is:

\[
h(n) = \frac{\Gamma(n+k)}{n!\Gamma(k)} \left( \frac{\mu}{\mu+k} \right)^k \left( 1 - \frac{\mu}{\mu+k} \right)^n
\]

Table 1 formulas defining discriminative and predictive performance indices.

| Sensitivity | Se(\( \lambda \)) = P(\( \Lambda > \lambda_i | N > \bar{n} \)) |
|-------------|--------------------------------------------------|
| Specificity | Sp(\( \lambda_i \)) = P(\( \Lambda \leq \lambda_i | N \leq \bar{n} \)) |
| PPV         | PPV(\( \lambda_i \)) = P(\( N > \bar{n} | \Lambda > \lambda_i \)) |
| NPV         | NPV(\( \lambda_i \)) = P(\( N \leq \bar{n} | \Lambda \leq \lambda_i \)) |
| Accuracy    | Acc(\( \lambda \)) = P(\( N > \bar{n}, \Lambda > \lambda_i \)) + P(\( N \leq \bar{n}, \Lambda \leq \lambda_i \)) |
| AUC         | AUC = \int_0^{\infty} Se(\( \lambda \)) dSp_{\lambda} (\( \lambda \)) d\( \lambda \) |

\( \lambda_i \) is a given cutoff value on the fall rate.
least twice ($\bar{n} = 1$). According to these two definitions, in the following we shall refer to 'prediction of any fall' or 'prediction of multiple falls', respectively.

### 3.2 Parameter Estimation and Results Visualization

We estimated the distribution $F$ for four different populations of community-dwelling and congregate-living older adults (Sydney, Melbourne, New Zealand and Atlanta). The data on fall counts were taken from [21]. Similarly to what already done in [21], for each population we fitted an intercept-only negative binomial regression, estimating the parameters $k$ and $\theta$.

Negative binomial regression was done with the function `glm.nb` of the package MASS of the R software [27]. We then derived the parameter $k$ and $\theta$ of the distribution of $\Lambda$ as $k' = k$ and $\theta = \frac{\mu}{\tau k}$.

MATLAB (R2011a) [28] has been used to plot and visually inspect the analytical formulas of Table 2. The AUC has been calculated via trapezoidal rule for numerical integration with the MATLAB function `trapz`.

### 4. Results

#### 4.1 Analytic Results: Performance Indices

Under the parametric assumptions presented in Methods, the joint probability of having a fall rate exceeding a given cutoff $\lambda_c$ and experiencing more than $\bar{n}$ falls is

$$P(A > \lambda_c, N > \bar{n}) = \frac{\gamma_U(\lambda_c/\theta, k) - (1 + \theta r)^{-k}\gamma_U(\lambda_c/\theta (1 + \theta r), k)}{1 - (1 + \theta r)^{-k}}$$

wheras the marginal probability of experiencing more than falls $\bar{n}$ is

$$P(N > \bar{n}) = \frac{\gamma_U(\lambda_c/\theta, k)}{\gamma_U(\lambda_c/\theta, k - 1)}$$

According to the definition given in Table 1, the sensitivity of the ideal prognostic tool is

$$Se(\lambda_c) = \frac{P(A > \lambda_c, N > \bar{n})}{P(N > \bar{n})}$$

The upper incomplete gamma function as $\gamma_U(x, k) = \frac{1}{\Gamma(k)} \int_0^x t^{k-1}e^{-t}dt$; $y_U(x, k) = 1 - \gamma_U(x, k)$; $\Gamma(k) = \int_0^{+\infty} t^{k-1}e^{-t}dt$.

### Table 2

Expected values of the discriminative and predictive performance indices of the ideal prognostic tool for falls, for the two definitions of faller.

<table>
<thead>
<tr>
<th>Fallers := ${\omega_i : N_i &gt; \bar{n}, \bar{n} = 0}$</th>
<th>Fallers := ${\omega_i : N_i &gt; \bar{n}, \bar{n} = 1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>$\gamma_U(\lambda_c/\theta, k) - (1 + \theta r)^{-k}\gamma_U(\lambda_c/\theta (1 + \theta r), k)$</td>
<td>$(1 + \theta r)^{-k}\gamma_U(\lambda_c/\theta, k) + \mu(1 + \theta r)^{-k-1}\gamma_U(\lambda_c/\theta (1 + \theta r), k + 1)$</td>
</tr>
<tr>
<td>Specificity</td>
<td>Specificity</td>
</tr>
<tr>
<td>$\gamma_U(\lambda_c/\theta, k)$</td>
<td>$(1 + \theta r)^{-k}\gamma_U(\lambda_c/\theta (1 + \theta r), k)$</td>
</tr>
<tr>
<td>PPV</td>
<td>PPV</td>
</tr>
<tr>
<td>$\gamma_U(\lambda_c/\theta, k) - (1 + \theta r)^{-k}\gamma_U(\lambda_c/\theta (1 + \theta r), k) - \mu(1 + \theta r)^{-k-1}\gamma_U(\lambda_c/\theta (1 + \theta r), k + 1)$</td>
<td>$\gamma_U(\lambda_c/\theta, k)$</td>
</tr>
<tr>
<td>NPV</td>
<td>NPV</td>
</tr>
<tr>
<td>$(1 + \theta r)^{-k}\gamma_U(\lambda_c/\theta, k)$</td>
<td>$\gamma_U(\lambda_c/\theta, k)$</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Accuracy</td>
</tr>
<tr>
<td>$\gamma_U(\lambda_c/\theta, k) + (1 + \theta r)^{-k}\gamma_U(\lambda_c/\theta (1 + \theta r), k)$</td>
<td>$\gamma_U(\lambda_c/\theta, k)$ + $\gamma_U(\lambda_c/\theta (1 + \theta r), k)$</td>
</tr>
<tr>
<td>$\gamma_U(\lambda_c/\theta, k) + (1 + \theta r)^{-k}\gamma_U(\lambda_c/\theta (1 + \theta r), k)$</td>
<td>$\gamma_U(\lambda_c/\theta, k)$ + $\gamma_U(\lambda_c/\theta (1 + \theta r), k)$</td>
</tr>
</tbody>
</table>

Notations:

$y_U(x, k) = \frac{1}{\Gamma(k)} \int_0^x t^{k-1}e^{-t}dt$; $y_U(x, k) = 1 - \gamma_U(x, k)$; $\Gamma(k) = \int_0^{+\infty} t^{k-1}e^{-t}dt$.
prognostic tool for falls can be derived similarly. Their expressions, reported in Table 2, have been obtained for both cases of prediction of any fall (\(\bar{n} = 0\)), and prediction of multiple falls (\(\bar{n} = 1\)).

### 4.2 Analytic Results: Accuracy Maximization

Following the definition given in Table 1, the accuracy can be computed as can be seen in Figure 2.

In order to look for a cutoff that maximizes the accuracy, we set its derivative to be zero (necessary condition). Equation 12 (Figure 3) is then determined by

\[
\lambda_{c,\text{max}} = \frac{\gamma_{U}(\lambda/c;\theta,k) - \gamma_{U}(\lambda/c;\theta(1+\theta\tau),k)}{1-(1+\theta\tau)^{-k}}
\]

The sensitivity of the performance indices with respect to the follow-up duration is shown in Figure 5, for the representative parameters extracted from the Sydney population. The AUC is only slightly affected, ranging from 0.84 at 8 weeks to 0.91 at 2 years. The PPV (NPV) is a monotonically increasing (decreasing) function of the follow-up duration. It ranges from 0.22 (0.98) at 8 weeks, to 0.88 (0.77) at 2 years.

Figure 6 shows the discriminative and predictive performances of the tool when prediction is made on any fall or multiple falls (\(\bar{n} = 0\) or \(\bar{n} = 1\), for the representative parameters extracted from the Sydney population. The AUCs for the four populations and the two predictions are: (Sydney, \(\bar{n} = 0\)) = 0.89, (Sydney, \(\bar{n} = 1\)) = 0.92; (Melbourne, \(\bar{n} = 0\)) = 0.82, (Melbourne, \(\bar{n} = 1\)) = 0.86; (New Zealand, \(\bar{n} = 0\))
= 0.85, (New Zealand, \( n = 1 \)) = 0.89; (Atlanta, \( n = 0 \)) = 0.80, (Atlanta, \( n = 1 \)) = 0.83.

### 5. Discussion

In this paper the predictive and discriminative performances of an ideal prognostic tool for falls have been evaluated by means of a probabilistic model. The indices considered for the evaluation have been the sensitivity, specificity, AUC, accuracy, and positive and negative predictive values. Although other metrics could have been considered (e.g. the Brier score and fractional reduction in entropy [13]), these have been chosen because they are by far the most commonly employed.

Having thus obtained the performances of a perfect prognostic tool for falls allows a critical assessment of some results that can be found when evaluating real prognostic tools. While it is known that, despite considerable research efforts, externally validated clinical tools still have modest performances [6, 10, 11], it is not infrequent that newly-developed tools come up with excellent but unlikely results. Having at hand some indicative reference values for the upper bounds of indices quantifying the goodness of the prediction is methodologically advisable and can suggest warnings against over-optimism.

The problem of over-optimism, often affecting newly-developed prognostic tools, has already been highlighted in the literature [8, 29]. A well-studied example is the STRATIFY (St Thomas Risk Assessment Tool in Elderly Inpatients). Without going into the details of the development and validation of this tool, it is worth mentioning that after being tested in several cohorts of older in-patients, a review [10] concluded that its prognostic performances are sensibly lower than previously reported by the first studies that led to its publication [30]. Another example could be represented by some recent sensor-based tools that have shown perfect accuracy [17].

One of the factors that may influence the reproducibility of prognostic tools (i.e. their capacity to keep their performance on subjects not included in the dataset used for original development, but similar for characteristics) is a low ratio between number of cases (number of fallers in our case) and number of candidate predicting variables available at the development stage (see the number of events per variable, EPV, discussed in [8]). This factor is critical in a context, like the development of sensor-based prognostic tools for falls, where (as yet) there is high availability of candidate variables. In this case, using statistical techniques that properly manage the high dimensionality of the problem (leveraging the so called ‘bias versus variance tradeoff’ [31]) and performing appropriately internal validation (e.g. cross-validation) are crucial. Furthermore, it is noteworthy that over-optimism may arise in the literature via publication bias even applying correct procedures of model fitting and validation. When the sample size is small, the estimation of the performance indices is subject to high variability, and studies with better results will be more likely published.

Finally, we point out that when the sample size is small, because of the variability on the estimated performance, even an imperfect prognostic tool for falls can outperform the limits here calculated. Reporting confidence intervals for the estimated parameters should hence be recommended in real applications. In our study the perfect tool was evaluated in the entire population, i.e. no sampling process has been modeled.

Among the analytic results, we have proven that the cutoff \( \lambda_{\text{max}} \) that maximizes the accuracy is independent of the population over which the prognostic model is evaluated. This result still holds if the accuracy is modified assigning different weights to true positives and true negatives (the two addends at the right side of Equation 10). Instead, it does not hold for other quantities (e.g. it does not hold for the Youden index). Clinical and economic considerations should lead to key choices for fall prevention strategies, as choices on the frequency of the assessment and the definition of faller (in terms of our notation, of \( r \) and \( n \)). Once these are made, our finding legitimates the practice of choosing a cutoff for a particular fall risk scale and applying it on different populations.

The populations in the Sydney, Melbourne and New Zealand studies are composed of community-dwellers aged 60, 70 and 80 years or more respectively. The population in the Atlanta study is composed of congregate-living, transitionally frail older adults aged 70 years or more. The diversity among these populations is reflected in the estimated parameters. As recalled in Methods, the mean of the fall rate distribution \( F \) is given by the product \( k b \). As expected, this quantity is highest for the Atlanta population (1.31 falls/year) and is lowest for Sydney (0.80 falls/year). In turn, this diversity is responsible of the heterogeneity in the values of the performance indices shown in Table 3 and Figure 4. As shown in Table 3, for a given cutoff, specificity is the parameter that varies most among the four populations, whereas PPV is quite stable. The AUC’s that we obtain are much higher than those found on validated clinical tools documented in the literature (see e.g. [11]). Thus, as expected, these traditional tools are far from providing a perfect probabilistic risk assessment.

Consistently over the four populations (results shown in Figure 5 only for the Sydney population), the AUC increases with the length of the follow-up as the net effect of the increase in specificity and decrease in sensitivity. The increase in specificity for a given cutoff can be explained in terms of its components: the proportion of true negatives (TN, subjects with fall rate less than the cutoff and never fallen during the follow-up) and the proportion of false

<table>
<thead>
<tr>
<th>Population</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Accuracy</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sydney</td>
<td>0.71</td>
<td>0.87</td>
<td>0.77</td>
<td>0.84</td>
<td>0.81</td>
<td>0.89</td>
</tr>
<tr>
<td>Melbourne</td>
<td>0.73</td>
<td>0.74</td>
<td>0.73</td>
<td>0.74</td>
<td>0.74</td>
<td>0.82</td>
</tr>
<tr>
<td>New Zealand</td>
<td>0.72</td>
<td>0.80</td>
<td>0.75</td>
<td>0.78</td>
<td>0.76</td>
<td>0.85</td>
</tr>
<tr>
<td>Atlanta</td>
<td>0.84</td>
<td>0.57</td>
<td>0.76</td>
<td>0.68</td>
<td>0.74</td>
<td>0.80</td>
</tr>
</tbody>
</table>
positives (FP, subjects with fall rate higher than the cutoff but never fallen during the follow-up). As the duration of the follow-up increases, more falls occur and both TN and FP decrease. However, as FP have higher expected fall rate, their decrease is quicker and this determines a net increase in specificity. For similar reasons, true positives (subjects with fall rate higher than the cutoff and fallen during the follow-up) and false negatives (subjects with fall rate less than the cutoff and fallen during the follow-up) are responsible for a decrease in sensitivity.

Consistently over the four populations, the AUC is slightly higher when the prediction is made on multiple falls (definition of faller with \( \bar{n} = 1 \)) than on any fall (\( \bar{n} = 0 \)), as the net effect of an increase in sensitivity and a decrease in specificity. Such results may indicate that one same tool is likely to show better discrimination when the prediction is made on multiple falls. This is consistent with what was found and commented in [32], although our findings about the predictability of multiple falls have been reached from a different perspective, without considering any knowledge other than the estimated distribution of the fall rate in the population. Indeed an accurate and accepted definition of who should be classified as a faller is still missing, and a matter of discussion in the literature. Lord and colleagues have proposed to define a faller as one fallen at least twice during the follow-up to filter out ‘occasional’ falls [3], i.e. with the objective to possibly identify falls which were due only to substantial and persistent risk factors, the only ones that can reasonably be predicted by a prognostic model.

Pointing at the distribution of fall rate in the population, the length of the follow-up, and the definition of faller as potential sources of heterogeneity for the reported performances of fall risk prognostic tools, this study links to the work of Haines et al [29], that explained part of the variability found in the literature in terms of differences in study design. Furthermore, in the present study we have showed how the effects of these factors on the discriminative and predictive performances of the tools rely on non-linear relations that standard models for meta-analysis cannot address.

All the results here obtained are valid within the hypotheses stated in Methods and the assumption that the expected fall rate of each subject is constant over time. However, the choice of a Poisson distribution for the conditional number of falls, given the fall rate, accommodates both the scenarios of time-constant and time-variable expected fall rate of each subject in the population, provided that the change in the expected fall rate is independent of the occurrence of a past fall. In particular, calling \( N_i(t) \) the random process representing the number of falls from baseline to time \( t \) of subject \( \omega_i \), the time-constant fall rate scenario is equivalent to assuming \( N_i(t) \) as a homogeneous Poisson process with intensity \( \lambda_i \), whereas the time-varying fall rate

**Figure 4** Performances of the ideal prognostic tool for different parameters \((k \text{ and } \theta)\) of the population fall rate distributions, follow-up length \( t = 1 \text{ year} \), \( \bar{n} = 0 \) (prediction on any fall). \( k \) and \( \theta \) parameterize the distribution of the fall rate \( \Lambda \) so that its mean is \( k\theta \) and its variance is \( k\theta^2 \). In panels a–c the performances of the ideal tool have been evaluated using the estimated shape and scale parameters \( k \) and \( \theta \) of the four populations: Sydney, Melbourne, New Zealand, Atlanta. The red dots mark the points corresponding to a cutoff value \( \lambda = \log(2) \text{ falls/year} \). Panels d, e: AUC and maximum accuracy (accuracy calculated for \( \lambda = \lambda_{c, \text{max}} \)) as functions of \( k \) and \( \theta \).
scenario is equivalent to assuming $N_i(t)$ as an inhomogeneous Poisson process with intensity function $\lambda_i(t)$, $\lambda_i$ being its mean from baseline to time $\tau$: $\lambda_i = \frac{1}{\tau} \int_0^\tau \lambda_i(u) du$. In this second scenario $\lambda_i$ is clearly dependent on the length of the follow-up. The sensitivity analysis with respect to the length of the follow-up, shown in Figure 5, is no more valid in this second scenario. Therefore, its results should be reconsidered if the expected fall rate of the subjects is believed to undergo substantial changes during the follow-up. In homogeneous and inhomogeneous Poisson processes the occurrence of an event is independent of the occurrence of any other. If this hypothesis is not valid and the change in the expected fall rate is supposed to be driven by the occurrence of a previous fall, other models should be considered (e.g. pure birth process) [33, 34]. However, it is worth noting that, given only the fall counts in a given time period, the identifiability among alternative models is not guaranteed [35]. Thus far, all the clinical tools have followed the approach of providing a unique score for the proneness to falling, without discerning for scenarios of subject-specific time-varying fall rate during the follow-up, nor has clinical epidemiology provided sufficient descriptive evidence for this kind of scenarios. Our choice to give a main focus to the case of constant fall rate has to be considered in this light and for the sake of simplicity.

The gamma distribution for $\Lambda$ was deduced from the hypothesis that the marginal distribution of the number of falls is of negative binomial type and from the identifiability of continuous mixtures of Poisson distributions [22, 23]. The hypothesis for this marginal distribution is supported from the empirical evidence shown in [21]. A gamma distribution for the fall rate has already been considered for negative binomial regression [36, 37]. The theoretical results about the identifiability of Poisson distributions makes the problem of estimating $F$ well posed. However, given a finite number of observations over the $N_i$’s, we cannot exclude that other distributions may fit equally well the data. A sensitivity analysis of the results presented in this paper with respect to other approaches for the estimation of $F$ will be considered for future investigations.

6. Conclusions

We have proposed the model of an ideal prognostic tool for falls which assumes that falls occur within a population according to the Greenwood and Yule scheme for accident proneness. We have derived analytically the performance indices of such perfect prognostic tool. We have estimated the parameters of the fall rate distribution of four different populations observed in different epidemiological studies, and we have then obtained quantitative evaluation of the analytical formulas. In the four considered populations, the AUC of the perfect tool, predicting any fall over a follow-up of one year, was estimated to range between 0.80 and 0.89.

We have shown that the performance indices of the perfect prognostic tool can
be estimated solely from falls counts and can be useful reference values for future works introducing new fall risk assessment tools. The analytical results give also an indication about how to choose a cutoff that maximizes the accuracy or any other weighted function of true positive and true negative rates. The maximum accuracy when prediction is made on any fall for a follow-up length \( \tau \) is attained with a cutoff of \( \log(2)/\tau \).

The model has allowed us to identify, analyze and quantify the effect of major factors that account for the high heterogeneity of results observed in the literature: i) the fall rate distribution over the population, ii) the length of the follow-up, and iii) the definition of faller as single faller or multiple faller. Because of the different fall rate distributions, specificity was found to have remarkable variations, varying over the four considered populations from 0.57 to 0.87. Predicting on multiple falls was found to have an effect on the AUC in terms of an increase of about 0.04 with respect to prediction on any fall.

**Acknowledgment**

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement FARSEEING n° 288940. Pierpaolo Palumbo is supported by a grant from Sovvenzione Globale Spinner 2013. The authors would like to thank Prof. Daniela Giovanna Calò for suggesting appropriate statistical references.

**References**
