Journal of Gerontology: MEDICAL SCIENCES
2005, Vol. 60A, No. 4, 530–534

Statistical Analysis of Efficacy in Falls Prevention Trials

M. Clare Robertson,[1] A. John Campbell,[1] and Peter Herbison[2]

[1]Department of Medical and Surgical Sciences, and
[2]Department of Preventive and Social Medicine, University of Otago Medical School, Dunedin, New Zealand.

Background. Many different and sometimes inappropriate statistical techniques have been used to analyze the results of randomized controlled trials of falls prevention programs for elderly people. This makes comparison of the efficacy of particular interventions difficult.

Methods. We used raw data from two randomized controlled trials of a home exercise program to compare the number of falls in the exercise and control groups during the trials. We developed two different survival analysis models (Andersen-Gill and marginal Cox regression) and a negative binomial regression model for each trial. These techniques a) allow for the fact that falls are frequent, recurrent events with a non-normal distribution; b) adjust for the follow-up time of individual participants; and c) allow the addition of covariates.

Results. In one trial, the three different statistical techniques gave surprisingly similar results for the efficacy of the intervention but, in a second trial, underlying assumptions were violated for the two Cox regression models. Negative binomial regression models were easier to use.

Conclusion. We recommend negative binomial regression models for evaluating the efficacy of falls prevention programs.

FALLS are common, recurrent events with serious consequences for elderly people. Randomized controlled trials of intervention programs have shown that these events can be prevented (1). However, there are no standard methods for evaluating and reporting the results from intervention trials that test programs to prevent falls; this lack of standard methods makes comparison of the efficacy of different interventions difficult.

A recent systematic review identified over 60 randomized controlled trials of interventions to prevent falls in elderly people, and these studies used a variety of approaches to test the efficacy of the intervention (1). Some were clinically and methodologically unsound.

Many studies reported recording all falls during the trial, but the results were presented using information on the first fall only. The clinical justification for a falls prevention program is to prevent as many falls as possible because all falls are likely to cause injury or have other adverse consequences. Using the time to a first fall or the proportion of fallers as the sole outcome ignores multiple events and wastes relevant information. For both clinical and statistical reasons, it is important to include all falls for each person as the outcome measure when evaluating the efficacy of a falls prevention program.

Some studies incorrectly assumed a normal distribution for falls and used Student's *t* test, linear regression, or analysis of variance. Survival analysis models make no assumption about the distribution of the outcome being compared, but there are other underlying assumptions that need to be tested, and the results of these tests were not often reported.

Intention-to-treat analysis requires that participants in randomized controlled trials are compared in the groups to which they were originally, randomly assigned. Even with the best attempts to minimize withdrawals, participants do die or withdraw during the trial; this is a particular issue in trials of elderly people. For some trials in the recent systematic review, the analysis was only for those completing the trial, whereas in others it remained unclear whether information from all participants was included (1). Dropouts in each group were usually reported, but whether dropouts resulted in the intervention or control groups being monitored for a significantly different length of time, and the possibility of bias resulting from this, was rarely discussed. There are statistical methods available for analyzing recurrent events which could take into account the individual follow-up time of participants in an intervention trial (2).

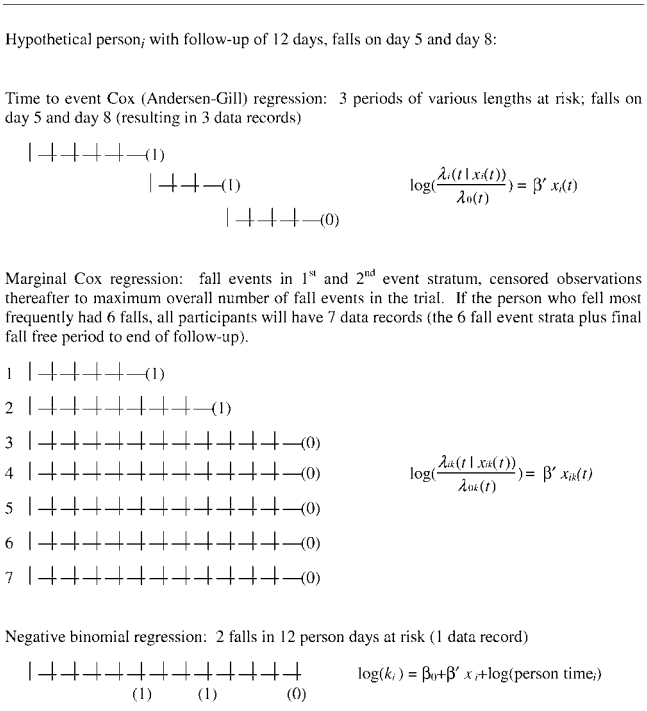
In this article we describe three different statistical modeling techniques in which all falls are considered, which allow for the fact that falls are frequent, recurrent events with a non-normal distribution over time, and which adjust for a) the variable follow-up time of participants during the trial and b) confounding variables, if appropriate. We used data from two of our four controlled trials of the Otago Exercise Programme to illustrate the use of the different techniques. We make a recommendation for future analyses of efficacy in falls prevention trials.

Methods

Overview

We used data on all falls at 1 year of follow-up collected in two randomized controlled trials of the Otago Exercise Programme, Dunedin Study A (3) and the West Auckland trial (4). The intervention, a set of leg-strengthening and balance-retraining exercises that increase in difficulty, plus a walking plan, was designed specifically to prevent falls in elderly people living in the community (5,6).

In both trials, participants were recruited through general practices and were randomly assigned to receive either the home exercise program or usual healthcare for 1 year.



*Adapted from Stürmer et al. (2).

Notes:
i (1,2,…., I) are participants, and k (1,2,…., K) fall events.

k_i: number of fall events of ith participant; (0): no fall event during or at the end of the interval, (1): event during or at the end of the interval; λ_i: hazard of a fall event for the ith participant; λ_{ik}: hazard of the kth fall event for the ith participant.

Figure 1. Schematic representation and statistical models of three different statistical techniques for evaluating recurrent events.*

Dunedin Study A (3), a trial of women aged 80 years or older ($n = 233$) in a research setting, was subsequently extended for another year (7), but the later results are not presented here. In Dunedin Study A, the control group also received social visits at home to match the home visits to exercise group participants by the physiotherapist delivering the exercise program (3). The West Auckland trial ($n = 240$) tested the home exercise program in men and women aged 75 years or older in a routine healthcare setting—a community home health service—and the program was delivered by a nurse who was trained and supervised by a physiotherapist (4).

Falls were the main outcome and were defined as ''unintentionally coming to rest on the ground, floor or other lower level'' (8). Fall events were monitored in all participants by asking them to return preaddressed, prepaid postcard calendars each month for 1 year.

For the present article, we analyzed the efficacy of the exercise program in each trial by using two survival-analysis techniques, Andersen-Gill (9) and marginal models (10), which are extensions of the Cox proportional hazards regression model. In these models, the dependent variable was time to each fall for each participant measured in days. The third technique was a negative binomial regression model where the dependent variable was the total number of falls for each participant during the trial (11). Group allocation was the independent variable in all the models, and each of the three techniques allowed the rate of falls in the exercise program compared with those in the control group to be expressed as a risk ratio. Robust estimates of variance were calculated for all models. A schematic representation of the three techniques is provided (see Figure 1).

The three techniques also allow for the addition of further covariates in the models where this is appropriate. Potential confounding variables were measured at baseline and follow-up in the two trials (3,4), but these did not add significantly to the models reported in this article.

Cox Proportional Hazards Regression

Proportional hazards or Cox regression models can test whether several factors (for example, intervention group, baseline prognostic factors) are independently related to the rate of a specific event. The models calculate the hazard function, which is the instantaneous likelihood of falling on any particular day. The estimate of the relative hazard for the variable in the model indicating group allocation is a ratio of the risk of falling in the intervention group to that in the control group throughout the trial.

Using survival probabilities to analyze time to fall events assumes that, at any time, participants who are censored before the end of the trial (withdraw from the trial) have the same risk of falling as those who complete the trial (12). An assumption of proportional hazards models is that the ratio of the risks of the events in the two groups is constant over time and that the ratio is the same for different subgroups of the data, such as age and sex groups (13). This is known as the proportionality of hazards assumption. No particular distribution is assumed for the event times, that is, the time from the trial start date for the individual to the outcome of interest (in this case, a fall event).

Andersen-Gill and marginal Cox proportional hazards regression.—The assumption required in traditional survival analysis that event times are independent is violated when there can be multiple events per person. The Andersen-Gill extension of the proportional hazards regression model and the marginal proportional hazards regression model are both recommended for analyzing these data (2,14). In both these techniques, the dependencies between failure times (fall events) are ignored, but robust estimates of variances are used to take into account the dependence of events.

For the Andersen-Gill and marginal Cox regression models that we developed, all trial participants were included in the analyses whether or not they withdrew from the exercise intervention during the trials. We censored data after 1 year or when the person withdrew completely from the trial.

The main difference between the two models is in the way that the risk sets are defined for each fall (15). The simplest method to implement is the Andersen-Gill, which follows a counting process approach and assumes that all falls are equal or indistinguishable. In contrast, the marginal model is based on the idea of marginal risk sets, and the data are treated as a competing risk data set, as if the fall events were unordered. Each event has its own stratum and each participant appears in all strata. Therefore for the two models, the data files for the analyses are set up in two different ways.

Table 1. Characteristics of Participants, Fall Events, and Follow-Up Times in Control and Exercise Groups in the Two Trials

	Dunedin Study A		West Auckland Trial	
Characteristic	Control Group ($N = 117$)	Exercise Group ($N = 116$)	Control Group ($N = 119$)	Exercise Group ($N = 121$)
Mean (*SD*) age, y	84.1 (3.4)	84.1 (3.1)	81.1 (4.5)	80.8 (3.8)
No. of women (%)	117 (100%)	116 (100%)	80 (67%)	80 (66%)
No. of fallers in previous year (%)	55 (47%)	47 (41%)	45 (38%)	44 (36%)
No. of falls during trial (%)	62 (53%)	53 (46%)	51 (43%)	38 (31%)
No. of fallers during trial	152	88	109	80
Falls per 100 person-years during trial	134.0	80.9	100.6	68.5
Number withdrawing before 1 y (%)*	7 (6.0%)	13 (11.2%)	21 (17.6%)	8 (6.6%)
Mean follow-up time (*SD*), months†	11.64 (1.65)	11.26 (2.31)	10.9 (2.7)	11.6 (1.9)
Total follow-up time, person-years	113.44	108.80	108.33	116.79

Notes: *Significant difference between control and exercise groups in West Auckland trial, $p = .010$.
†Significant difference between control and exercise groups in West Auckland trial $p = .028$.

For the Andersen-Gill models the data records for each participant consisted of a separate record for each of that person's fall events plus a final record containing the time in days from the last fall until withdrawal from the trial or the end of the trial (15). If the person had no falls, there was one data record only which included the number of days the person was monitored.

The data files for the marginal models required, for each participant, a separate record for each fall for the maximum overall number of fall events in the trial (the most frequent faller in Dunedin Study A had 10 falls and in the West Auckland trial had 15 falls) plus a final record containing the total time of follow-up for that person (15). This resulted in 2563 and 3840 data records for the two trials, respectively.

We tested the Andersen-Gill and marginal models to determine whether they met the assumption that the hazard ratio is proportional over time. The test we used is equivalent to testing that the log hazard ratio function is constant over time; the result is a chi-square value which indicates deviation from the assumption (11).

A log likelihood value is associated with each Andersen-Gill model and each marginal model. If this value is close to zero, the model is considered to describe the data well, that is, it is a ''good'' model. When comparing models using the same data, the model best fitting the data is indicated by the log likelihood value nearer to zero. We compared the goodness-of-fit of the Andersen-Gill model with that of the marginal model developed from the same data by noting the model log likelihood value nearer to zero.

Negative Binomial Regression

The negative binomial regression model can also be used to compare recurrent event rates in different groups; it allows investigation of the treatment effect and confounding variables, and adjusts for variable follow-up times by using time at risk as the offset (16). These models are a generalization of the more commonly used Poisson regression model.

Poisson distributions assume that a) the recurrent events being counted are occurring independently of each other and randomly in time and b) the mean and variance are equal (13). Overdispersion is the term used when, as is the case with falls, there is more variation than would be expected in a Poisson process, that is, the variance of the count variable is greater than the mean (11). Thus overdispersion models attempt to take into account the dependence of events in the same individual. One alternative is to estimate a factor to correct the inferential statistics of the Poisson regression model; another alternative is to use the negative binomial regression model. Poisson is a special case of the negative binomial and corresponds to an overdispersion parameter of zero. The data files for the negative binomial regression analyses had one record per person containing simply the total number of falls during the trial and the total number of days the person was monitored (233 and 240 data records for the two trials, respectively). In our negative binomial models we noted whether the data best fitted a Poisson or a negative binomial model by testing whether the over-dispersion parameter equalled zero [indicated by the p value associated with a likelihood ratio chi-square (11)].

Table 2. Relative Ratio Estimates for All Falls in the Two Trials

Estimate	Dunedin Study A	West Auckland Trial
Relative hazard (95% confidence interval [CI]*)	0.60 (0.42–0.87)	Evidence that underlying assumptions were violated for this model
Log likelihood value	−1290.16	
Relative hazard (95% CI)†	0.56 (0.36–0.86)	Evidence that underlying assumptions were violated for this model
Log likelihood value	−1243.65	
Incidence rate ratio (95% CI)‡	0.61 (0.42–0.89)	0.54 (0.32–0.90)

Notes: *From Andersen-Gill extension of Cox regression model.
†From marginal Cox regression model.
‡From negative binomial regression model.

Results

Characteristics of trial participants, fall events, and follow-up times for control and exercise program groups in the two trials are presented in Table 1, and the risk ratio estimates are presented in Table 2. In the West Auckland trial, more participants from the control group withdrew before the end of the trial, resulting in exercise group participants being monitored for a significantly longer time (Table 1). This may have resulted from the lack of regular personal contact for the control group, as exercise group participants had monthly visits or telephone calls from the exercise instructor.

In Dunedin Study A the three statistical approaches indicated that falls were significantly reduced by 40% (Andersen-Gill Cox model), 44% (marginal Cox model),

and 39% (negative binomial regression model) in the exercise group compared with those in the control group. The tests for the proportionality of hazards for both types of survival regression models indicated that this assumption held for these models. The marginal regression model gave a slightly wider 95% confidence interval for the relative hazard than did the Andersen-Gill model, but the log likelihood values for the two models were similar (see Table 2).

In the West Auckland trial data there was evidence that the proportional hazards assumption was violated in the Andersen-Gill and marginal Cox regression models (proportional hazards test, group $\rho = 0.084$, $\chi^2 = 3.88$, $df = 1$, $p = .049$, and proportional hazards test, group $\rho = 0.079$, $\chi^2 = 4.09$, $df = 1$, $p = .043$, respectively). Therefore, a single hazard ratio describing the effect of the exercise program would be inappropriate.

In the West Auckland trial the negative binomial model indicated there was a 46% reduction in the number of falls during the trial for the exercise group compared with the control group (incidence rate ratio 0.54, 95% confidence interval 0.32–0.90). The likelihood ratio test of the overdispersion parameter for both the negative binomial models indicated that these models fitted the data significantly better than a Poisson model ($\chi^2 = 107.91$, $df = 1$, $p = .0001$ for Dunedin Study A; $\chi^2 = 143.03$, $df = 1$, $p = .0001$ for West Auckland trial).

Discussion

We have demonstrated the use of three statistical modeling techniques which use all falls as the outcome measure and are, therefore, appropriate for analyzing the efficacy of an intervention aimed at preventing falls. Different approaches to handling the problem of missing data are appropriate in different situations (17). When all participants cannot be followed to the end of a falls prevention trial, a good approach is to make use of the survival analysis and negative binomial regression models described. These models use data right up until the time of withdrawal or trial completion, and take into account each individual's time in the trial. They also take into account the fact that falls are recurrent events with a Poisson type distribution and attempt to adjust for the nonindependence of fall events in any one person.

Therneau and Hamilton (14) state that, when using survival analysis to analyze recurrent events, the most difficult to analyze are those in which the events are of the same type. Survival analysis is not valid if participants who are censored (withdraw before the end of the trial) do not have the same rate of outcome (risk of falling) as those who continue in the trial (18). Those not completing a falls prevention trial are more likely to be at higher risk of falling (7) and, if fewer from one group than another group withdraw, results may be biased. As the Andersen-Gill extension of the Cox model appears to be least affected by this problem, Therneau and Hamilton suggest it should be the first choice. They also favor the marginal Cox model if the proportional hazards assumption ''does not show great departures.'' The negative binomial regression model has been recommended for analysis of counts and rates in psychology (16), for motor vehicle crashes (19), and for other right-skewed data with clumping at zero (20). Several falls prevention intervention trials which used this model for analysis of efficacy have now been reported (4,21–23). The incidence rate ratios calculated in each model are easy to interpret, and this technique is now available in standard software (11).

There is no one statistical technique which is the ''right'' one for testing efficacy of falls prevention interventions, but some techniques have advantages over others. We found that using the negative binomial regression model for evaluating efficacy of the intervention in our falls prevention trials was simple, straightforward, and had several advantages over survival analysis techniques. Underlying assumptions were violated in some of the Cox regression models. In contrast to the complicated data files needed for the survival analysis models, the negative binomial regression analysis required only one data record per participant.

We found negative binomial regression models useful when analyzing efficacy in a controlled trial and in a meta-analysis of the individual level data from the four controlled trials of the Otago Exercise Programme (21,24). In the Southern New Zealand trial (21), centers rather than individuals were allocated to receive the exercise program or to act as controls, so that clustering on center was required in the analysis. In the meta-analysis, the main outcome was injuries resulting from falls; we clustered on the trial center and also added interaction terms to the models for subgroup analyses (24).

A number of successful falls prevention trials have now been reported (1). All interventions require healthcare resources, and healthcare planners need to be able to compare successful programs. Robust analytical techniques, using data for all falls, are essential for this. Similar techniques would certainly make comparison easier and would also enable pooling of data. We recommend using negative binomial regression models for evaluating the efficacy of falls prevention programs.

Acknowledgments

We thank biostatistician Sheila Williams, Department of Preventive and Social Medicine, University of Otago Medical School. The two falls prevention trials were funded by the Accident Compensation Corporation and the Health Funding Authority Northern Division, New Zealand.

Address correspondence to Dr. M. Clare Robertson, Department of Medical and Surgical Sciences, University of Otago Medical School, P.O. Box 913, Dunedin, New Zealand. E-mail: clare.robertson@stonebow.otago.ac.nz

References

1. Gillespie LD, Gillespie WJ, Robertson MC, Lamb SE, Cumming RG, Rowe BH. *Interventions for preventing falls in elderly people (Cochrane review)*. The Cochrane Library, Issue 3, 2003. Oxford: Update Software; 2003.
2. Stürmer T, Glynn RJ, Kliebsch U, Brenner H. Analytic strategies for recurrent events in epidemiologic studies: background and application to hospitalization risk in the elderly. *J Clin Epidemiol*. 2000;53:57–64.
3. Campbell AJ, Robertson MC, Gardner MM, Norton RN, Tilyard MW, Buchner DM. Randomised controlled trial of a general practice programme of home based exercise to prevent falls in elderly women. *BMJ*. 1997;315:1065–1069.
4. Robertson MC, Devlin N, Gardner MM, Campbell AJ. Effectiveness and economic evaluation of a nurse delivered home exercise programme to prevent falls. 1: Randomised controlled trial. *BMJ*. 2001;322:697–701.

5. Accident Compensation Corporation. Otago Exercise Programme to prevent falls in older adults. August 2003. Available at: http://www.acc.co.nz/injury-prevention/growing-and-living-safely/. Accessed February 8, 2005.
6. Gardner MM, Buchner DM, Robertson MC, Campbell AJ. Practical implementation of an exercise-based falls prevention programme. *Age Ageing*. 2001;30:77–83.
7. Campbell AJ, Robertson MC, Gardner MM, Norton RN, Buchner DM. Falls prevention over 2 years: a randomized controlled trial in women 80 years and older. *Age Ageing*. 1999;28:513–518.
8. Buchner DM, Hornbrook MC, Kutner NG, et al. Development of the common data base for the FICSIT trials. *J Am Geriatr Soc*. 1993;41:297–308.
9. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann Stat*. 1982;10:1100–1120.
10. Lin DY. Cox regression analysis of multivariate failure time: the marginal approach. *Stat Med*. 1994;13:2233–2247.
11. StataCorp. *Stata Statistical Software: Release 6.0*. College Station, TX: Stata Corporation; 1999.
12. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). *BMJ*. 1998;317:1572.
13. Kirkwood BE. *Essentials of Medical Statistics*. Oxford, United Kingdom: Blackwell Science Ltd.; 1988.
14. Therneau TM, Hamilton SA. rhDNase as an example of recurrent event analysis. *Stat Med*. 1997;16:2029–2047.
15. Cleves M. Analysis of multiple failure-time data with Stata. *Stata Technical Bulletin* 1999;49:30–39.
16. Gardner W, Mulvey EP, Shaw EC. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychol Bull*. 1995;118:392–404.
17. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*. 1999;319:670–674.
18. Katz MH, Hauck WW. Proportional hazards (Cox) regression. *J Gen Intern Med*. 1993;8:702–711.
19. Milton J, Mannering F. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation*. 1998;25:395–413.
20. Byers AL, Allore H, Gill TM, Peduzzi NP. Application of negative binomial modeling for discrete outcomes. A case study in aging research. *J Clin Epidemiol*. 2003;56:559–564.
21. Robertson MC, Gardner MM, Devlin N, McGee R, Campbell AJ. Effectiveness and economic evaluation of a nurse delivered home exercise programme to prevent falls. 2: Controlled trial in multiple centres. *BMJ*. 2001;322:701–704.
22. Nikolaus T, Bach M. Preventing falls in community-dwelling frail older people using a home intervention team (HIT): results from the randomized falls-HIT trial. *J Am Geriatr Soc*. 2003;51:300–305.
23. Barnett A, Smith B, Lord SR, Williams M, Baumand A. Community-based group exercise improves balance and reduces falls in at-risk older people: a randomised controlled trial. *Age Ageing*. 2003;32:407–414.
24. Robertson MC, Campbell AJ, Gardner MM, Devlin N. Preventing injuries in older people by preventing falls: a meta-analysis of individual-level data. *J Am Geriatr Soc*. 2002;50:905–911.

Received September 3, 2003
Accepted October 13, 2003
Decision Editor: John E. Morley, MB, BCh