Appendix I
References for all articles included in the systematic review (n=83)


50. McMurdo ME, Mole PA, Paterson CR. Controlled trial of weight bearing exercise in older women in relation to bone density and falls.[see comment]. *BMJ*. Feb 22 1997;314(7080):569.


Appendix II: Overview of 3 statistical methods to analyse recurrent events

The rate of falls is most often modelled using Poisson regression. Poisson regression assumes that events are independent, the mean and variance are equal in the population of interest and that the event rate is constant over time. These assumptions may not hold for recurrent events. When the event rates are compared using Poisson regression without accounting for the dependence of events within individuals the confidence intervals may be too narrow and the p-value too small -- thus a 'truly ineffective' treatment might be reported as effective (Type I error). Several statistical models do not require these assumptions, including negative binomial regression, the Andersen-Gill extension of the Cox model and the Wei Lin and Weissfield marginal model. These, and other methods, have been described in detail in a recent publication by Kuramoto and colleagues. This article describes the type of research question addressed in each method, interpretation of the rate ratios, data set structure and SAS code. Similar to a study by Robertson and colleagues, this paper uses a simulation study to compare several different statistical methods and interpret their rate ratios.

Under a Poisson process, the counts of recurrent events are Poisson variants but it is often the case that they display substantial extra-Poisson variation (overdispersion)—that is, the variance is greater than the mean. Overdispersion is usually attributed to, a) the dependence of events within participants or b) the
dependence of events over time when previous events alter the risk of subsequent events. One approach that can be used to model the dependence within individuals is to introduce a random effect or ‘frailty’ variable $\alpha_i$ for each subject. The $\alpha_i$'s are independent and identically distributed random variables with some distribution function $G(\alpha)$. Such models are referred to as random-effects, mixed-effects or frailty models. When the frailties are distributed according to a Gamma distribution the actual counts of falls then follow a negative binomial distribution. The negative binomial distribution model assumes that the event rate is constant over the intervention period.

Random-effects (or frailty) models are an improvement over Poisson models--the point estimate is the same but the standard error of the estimate is increased, therefore widening the confidence intervals. However, they are not a perfect solution as these models ignore the timing of the events and cannot compensate for events clustering in the population (e.g., during an episode of acute illness such as the flu) or variations in the rate of falls over the natural course of time (e.g., seasonal variation, secular trends). For these reasons, there is utility in recurrent event analyses that consider the total number of events and their timing-- including the Andersen-Gill model and the Wei, Lin and Weissfield (WLW) marginal model.

The Andersen-Gill model, an extension of the Cox proportional hazards model, does not require the assumption of constant rate, or even to specify the baseline rate. The Andersen-Gill model, however, does not account for the order of the events and therefore the relationship between previous falls and subsequent falls is not dealt with explicitly. This may be crucial in falls prevention studies as a person who falls once is more likely to fall again. The regression coefficient(s) (similar to the negative binomial model) estimated by the Andersen-Gill model demonstrate the treatment effect when falls (or the events of interest) within subjects are studied together without accounting explicitly for the order of events. For example, the rate estimated by the model would be the same, for example, 55% reduction be it the 1st, 2nd, 3rd or greater fall.
The Wei, Lin and Weissfield (WLW) marginal model assumes that an individual is at risk to experience all possible events to the maximum number of events observed in the sample. For example, if the maximum number of falls observed is 10, then all participants in the WLW model are considered at risk for 10 falls even though they may not have fallen at all. The WLW model allows both the fall-specific effect of treatment to be estimated as well as the pooled effect of treatment (i.e. the average effect of the treatment over all events). The regression coefficients for the pooled and fall-specific effect of treatment can be stated as follows: The overall risk of a fall is 49% lower in the treatment group compared to the control group. The risk of a 1st, 2nd and 3rd fall is 49, 25, and 10% lower respectively in the treatment group compared with the control group. Interestingly, the fall-specific estimates can be used to test the null hypothesis that the fall-specific regression coefficients are equal. If the null hypothesis is rejected we assume that the alternative hypothesis is true and that one or more of these coefficients are different.

Estimating fall-specific coefficients may be beneficial in ascertaining the impact of the intervention of the risk for future events. Although the WLW model produces fall-specific coefficients out to the maximum number of events experienced in the sample, the interpretation of these coefficients, particularly the 3rd, 4th and later events, is difficult as there may be very few participants to whom these may apply. These methods are summarized in Table 1.

The Mean Cumulative Function (MCF) is an additional method used in recurrent event analysis which estimates the average number of events per participant over time with 95% confidence interval and the results can be presented graphically. The MCF difference can also be calculated with its corresponding 95% confidence interval. The MCF is illustrated in Figure 1, using two groups labeled A (a control group) and E (an intervention group). The MCF difference (comparing Group A to Group E) is illustrated in Figure 2. In Figure 1, the curves begin to separate at approximately day 50, suggesting that this is when the intervention begins to take effect. In Figure 2, at approximately 250 days the intervention (Group E) is, on
average, preventing one event per person compared with the control group (Group A). A recent publication has used the MCF to compare several simulated intervention scenarios.

Sample size can potentially be reduced when all events are considered in the analysis compared with time to first event analysis. Cook uses the example of a trial with 90 percent power with an effect size of 0.50 with a two-sided test at alpha=0.05. When the recurrence of events is considered in the sample size calculation, it was estimated that the trial duration would be 2.71 years and 163 patients would need to be recruited. However, if the same trial is based on time to first event then the trial duration would be 2.97 years and 178 patients would need to be recruited. However, these calculations are not straightforward and factors such as subject accrual, duration of follow-up, event rate parameters and dropout rates must be considered. An extensive discussion of these issues is given by R.J. Cook. Consultation with a biostatistician is highly recommended.
Table 1 Assumptions, strengths and limitations of three recurrent event statistical methods

<table>
<thead>
<tr>
<th>METHOD</th>
<th>ASSUMPTION(S)</th>
<th>STRENGTH(S)</th>
<th>OUTPUT</th>
<th>LIMITATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEGATIVE BINOMIAL REGRESSION= Mixed EFFECTS POISSON MODEL= RANDOM EFFECT (FRAILTY) MODEL ASSUMING A GAMMA DISTRIBUTION</td>
<td>SURVIVAL TIME IS UNRELATED TO EVENT RATE</td>
<td>ACCOMMODATES VARIABLE FOLLOW-UP TIMES AND CAN ADJUST FOR OTHER STUDY FACTORS</td>
<td>INCIDENCE RATE RATIO</td>
<td>DOES NOT CONSIDER THE TIMING OF EVENTS</td>
</tr>
<tr>
<td></td>
<td>EVENT RATE IS CONSTANT</td>
<td>CAN INCORPORATE ROBUST STANDARD ERRORS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CAN ADJUST FOR CLUSTERS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANDERSEN-GILL MODEL= INDEPENDENT INCREMENT MODEL</td>
<td>ALL FALLS (EVENTS) CONSIDERED EQUAL</td>
<td>ACCOMMODATES VARIABLE FOLLOW-UP TIMES AND CAN ADJUST FOR OTHER STUDY FACTORS</td>
<td>HAZARD RATIO</td>
<td>ORDER OF EVENTS NOT EXPLICITLY HANDLED</td>
</tr>
<tr>
<td>WLW MARGINAL MODEL</td>
<td>ALL PARTICIPANTS CONSIDERED ‘AT RISK’ FOR ALL FUTURE EVENTS</td>
<td>ACCOMMODATES VARIABLE FOLLOW-UP TIMES AND CAN ADJUST FOR OTHER STUDY FACTORS FALL-SPECIFIC (EVENT SPECIFIC) AND POOLED TREATMENT EFFECTS ESTIMATED</td>
<td>HAZARD RATIO INTERPRETATION OF COEFFICIENTS FOR LATER EVENTS CHALLENGING AS VERY FEW PARTICIPANTS MAY EXPERIENCE LATER EVENTS</td>
<td>CAN INCORPORATE A RANDOM EFFECT OR ROBUST STANDARD ERRORS CAN ADJUST FOR CLUSTERS 5</td>
</tr>
</tbody>
</table>
Figure 1. The mean cumulative function for Group A and Group E
Figure 2. The mean cumulative function difference for Group A compared to Group E
References


