Methods for analysing recurrent events in health care data. Examples from admissions in Ebeltoft Health Promotion Project

Janus L Thomsen\textsuperscript{a} and Erik T Parner\textsuperscript{b}


Background. Evaluation of health care contacts from first events alone often misses large amounts of potentially important data and may produce different results than evaluation of all data including recurrent events.

Objective. We aim to bring the different methodological approaches for analysing longitudinal health care data to the attention of researchers in primary care.

Methods. We used hospital admission data from the Ebeltoft Health Promotion Project, a randomized trial in primary care examining the effect of preventive health checks. Comparisons included three randomized groups: an intervention group receiving health checks, a group where intervention consisted of a health check followed by a health discussion with the GP and one control group.

Results. Both intervention groups had \(
\approx 20\% \) fewer hospital admissions than the control group over a 6 year period. If dependence among recurrent events is excluded, such a reduction amounts to a highly significant effect. Use of the standard Poisson distribution for analysing recurrent events and exclusion of their dependent structure causes data interpretation to be incorrect, because the model does not account for the extra variability between persons; the resulting 95\% CIs would therefore be too small.

Conclusion. Analysis of health care contacts should embrace both first and recurrent events and it should use a model appropriate to these data. An individual rate model that includes a parameter of an unspecified individual event distribution frailty may be a natural choice when analysing longitudinal data of contacts to the health care system in broad terms.

Keywords. Poisson distribution, recurrence, statistics, epidemiological methods, health care utilization.

Introduction

Data comprising multiple events per person, like number of health care contacts over time compared between groups of individuals, are common in both clinical and epidemiological studies. A natural measure of the number of events is the ‘event rate’. One of the main methodological challenges is how to address the dependence of multiple events.

For example, it may empirically be expected that during a fixed time interval some persons will have no admissions, some will have one or two, while a small number may have many. Such persons obviously are not facing the same risk and admissions within individuals cannot be seen as independent events.

Although well established in the biostatistics literature\textsuperscript{1} and to some extent in the epidemiological literature,\textsuperscript{2} methods for overcoming these challenges are rarely encountered in the primary health care literature and only briefly mentioned in standard textbooks of epidemiology.\textsuperscript{3}

This article discusses different methodological approaches for analysing longitudinal data using hospital admission data from the Ebeltoft Health Promotion Project.\textsuperscript{4} Comparisons included three randomized groups: an intervention group receiving health checks.
(Intervention 1), a group where intervention consisted of a health check followed by a health discussion with the GP (Intervention 2) and one control group. Both intervention groups had ~20% fewer hospital admissions than the control group over a 6 year period.4

A simple analysis of the rates in the three groups based on the Poisson distribution gave narrow confidence intervals (CIs) around the point estimates of ~20% fewer admissions in the intervention groups, indicating a clear effect of the interventions (P < 0.001). This paper explains why this Poisson analysis is inappropriate and it presents a method for analysing the recurrent events that captures the individual proneness for admissions by a so-called frailty variable.

Data from the Ebeltoft Health Promotion Project and the study design are described in detail in previous publications.5 Data were analysed using Stata Statistical Software: Release 7.0 (Stata Corporation 1999).

First event or multiple events

Analysis of rates of recurrent events are often handled with statistical approaches only appropriate for analysis of first event and it has therefore been suggested to replace analysis of repeated events with analysis of first events.6

Whether either approach is appropriate evidently depends on the scientific question posed. Rates of first events may be appropriate if such events affect study adherence or other behaviour and risk factors for disease, for example trials investigating diagnoses like specific cancers or stroke. Consideration of all events may be more relevant in other settings where events are more frequent and some patients studied may already have had previous events. In such situations restriction of analysis to first event would discard much relevant information and may be difficult in settings where patients have previous events.

The literature holds several examples of such reduction of data to first event.6–10 Glynn et al.8 thus illustrate the importance of investigating multiple events with data from a clinical trial of the effect of intervention in the form of regular intake of cranberry juice drink on bacteriuria in elderly women. After randomization, six urine samples were collected at roughly monthly intervals.10 The cranberry group had 18% fewer first events than the placebo group. This could be explained by coincidence in which case the study would have concluded that there was no statistically significant difference between the groups.8

However, analysis of all urine specimens collected throughout the study disclosed that 50% fewer of the urine samples in the cranberry group were positive compared with the placebo group. This difference in recurrent bacteriuria was highly statistically significant.8 The discrepancy between rates of first events and overall rates arose because women in the cranberry group were more likely in the long term to recover than women in the placebo group, and restriction of analyses to only first events would hence have obscured the effect of cranberry juice.

Analysis of time to first health care contact is not relevant when using health care utilization as a proxy for a population’s morbidity because it omits much information about its total morbidity. Analysis of time to first event may also give misleading results in settings where intervention may cause admission rates to rise. Consideration of all admissions can accordingly change the estimates drastically.

The use of inappropriate statistical methods for analysis of the rate of recurrent events may be rooted in unfamiliarity with both the Poisson distribution and the extended Poisson models. This article therefore offers a review of the arguments for using the Poisson distribution for analysing longitudinal data.

First event

Models for the relative risk and incidence rate ratio for first event

In longitudinal data a risk of an event generally refers to the probability of that event occurring within a certain amount of time. If all subjects in a group are followed for the same period of time and the event only occurs one time per subject, the cumulative incidence proportion (CIP) is calculated as the number of subjects who experience the event divided by the total number of subjects in the group. The relative risk (RR) compares two independent CIPs. Statistical analysis of the CIP and the RR rests on a binomial data model, because an event either occurs or does not occur during the period under consideration. Table 1 shows the number of admissions and the time at risk with CIPs and RRs calculated by the binomial model risk for the three study groups in the Ebeltoft Health Promotion Study.

Where all subjects cannot be followed over the same time-period, the standard procedure is to calculate the incidence rate which is defined as the number of events in a group divided by the sum of individual time at risk for subjects in the group, when the subjects are no longer at risk after the event (Fig. 1, upper panel). The incidence rate ratio (IRR) compares two incidence rates and was virtually identical to the RR for hospital admission (Table 2), because the individuals in the groups were followed for approximately the same amount of time.

The basic model usually associated with rates is the Poisson distribution, like the basic model for proportions is the binomial distribution and the basic model for the mean is the normal distribution. If we observe n individuals in a study with the probability p for one
of two possible outcomes, the binomial model $\text{Bi}(n, p)$ assumes that the number of individuals $n$ is fixed and therefore does not depend on the sampling, i.e. a previous outcome in the same study. Furthermore, the probability of the event has to be the same for all individuals and the observations in the $n$ individuals must be independent.

The argument for using the Poisson model as the basic model for rates in the longitudinal setting lies in its approximation to the binomial distribution known as the Law of Small Numbers. In its simple form the results show that if the sample size $n$ is large and the probability of event $p$ is small, then the binomial model and the Poisson model are almost identical. The expected number of events in the former, $n \times p$, is therefore equal to the expected number in the latter. The expected number in the Poisson distribution is often written as $T \times \lambda$, where $T$ is the sum of the observed time at risk of the individuals and $\lambda$ is the expected number of events per time unit observed. The Poisson approximation to the binomial distribution is good if the sample size is large. For small datasets as a rule of thumb, a good approximation is obtained when the sample size is above 20 and the probability of an event is below 5%. A very good approximation is obtained if the sample size is above 100 and the probability of an event is below 10%.

Setting the observed number of events $x$ equal to the expected number of events $T \times \lambda$, we obtain the well-known estimate of the rate (the unknown parameter) as the number of events divided by the sum of the observed time at risk, $x/T$. This estimate can also be derived by the more general method of maximum likelihood estimation. The likelihood function generally describes how the probability of the observed data (the likelihood) depends on the unknown parameter value. The maximum likelihood estimate is the parameter value that maximizes the likelihood of the observed data. The method of maximum likelihood is used both to construct estimates of the unknown parameters and to derive their statistical properties.

Rates are, however, used more generally than for analysing rare events. In the Poisson model we focus on the number of events which is set in relation to the cumulative time at risk. We may, however, also view the data as time to events as is done in survival analysis. In survival analysis the rate (hazard rate) refers to the risk of an event during a very brief period, assuming no event has occurred at the beginning of the period, and it can be interpreted as the risk of an event at a specific time. The hazard rate describes the rate in a very short interval of time as a function of time. The simplest scenario is when the hazard rate is constant. This model is termed the exponential model and it is closely connected to the Poisson model. One important similarity is that the likelihood functions in the two models are identical. This implies that the estimates of the rate derived in the two models and their statistical properties are identical. The two models are only approximately the same when the events are rare. When analysing rates for longitudinal event data, the exponential survival model is generally the underlying statistical model. Many extensions of rates to recurrent events can be formulated both in extension of the Poisson model and of the exponential survival model.
Standard Poisson regression

Simple comparisons of rates between two groups of subjects may not be valid because they may differ on important confounding variables. A common strategy in the presence of confounding is to present standardized rates, which are composed of weighted stratum-specific rates, where the strata are formed by categories of the confounding variables. Regression approaches, however, will often be more efficient when several confounders have to be controlled for simultaneously. The results of the crude regression analysis of first events without controlling for confounding variables are shown in Table 2. Individuals with events before the study period are not excluded.

Stratification of rate ratios for hospital admissions at annual intervals in the Ebeltoft Health Promotion Study revealed an interesting trend with a significant fall 4 and 5 years after intervention launch.4 Rate ratios for hospital admissions in the intervention groups compared with the control group considering the entire follow-up period did not change when adjusting for time as a categorical variable (data not shown).

If the rate is not constant over time, we may include time as a categorical analytical variable. Time-dependent rates can also be handled within a Cox regression analysis. This will theoretically be approximately the same as dividing time into very small intervals in a Poisson regression analysis. Results for Cox regression of first event are presented in Table 2. Methods for Cox regression on recurrent events do exist, but for simplicity we focus on extensions of the Poisson model.

Recall that event rates are expressed per person-years. Rate ratios are usually presented for comparisons within groups, and are calculated as ratios of the rates (Table 2).

Recurrent events

IRRs of recurrent events

When patients may have had previous events and more such events are considered, the analysis becomes fundamentally different from analysis of first events (Fig. 1, lower panel). When more events are studied for each individual, the models have to account for the extra variability between persons, as some persons will tend to have experienced many and other only few or no events.

In the Ebeltoft Health Promotion data, both intervention groups had ~20% fewer hospital admissions than the control group over a 6 year period. If dependence among recurrent events is excluded, such a reduction amounts to a highly significant effect: Rate ratios for intervention compared with control was 0.78 (95% CI: 0.69–0.89, P < 0.001) for group 1 and 0.81 (95% CI: 0.71–0.92, P < 0.001) for group 2. Use of the standard Poisson distribution for analysing recurrent events and exclusion of their dependent structure causes data interpretation to be incorrect, because the model does not account for the extra variability between persons; the 95% CIs and P-values would therefore usually become too small.

Rates and rate ratios for recurrent events are shown in Table 3, where it is seen that there was no difference in rate ratio estimates between the first and recurrent events. This is, however, not always the case and, as mentioned above, the literature holds several examples of reduction of data to first events, which conceals information.6–10 The analysis in Table 3 is described in the following.

Regression models with recurrent events

An important dimension of recurrent events is that the variability is greater than expected by the standard Poisson distribution, which is often referred to as over-dispersion. Figure 2 illustrates the different tendencies of events (frailties) to occur among individuals; the dots mark the individual rates, the dotted line the average rate in the group. The individual frailty variable Zt may be illustrated as the rate ratio between the individual rates compared with the average rate in the group.

The negative binomial distribution assumes that the frailty variable follows a Gamma distribution with mean one and variance \( \theta \).1 The parameter \( \theta \) describes the heterogeneity between individuals. A large \( \theta \) value indicates much heterogeneity between individuals and a small \( \theta \) value less heterogeneity. If \( \theta \) is very small, the negative binomial distribution is approximately equal to the Poisson distribution, i.e. the risk of admissions is approximately the same for all individuals. This model belongs to a group of models called...
Neg.Bi: Po($Z \cdot \lambda \cdot T$), where $Z$ follows a gamma distribution with mean $= 1$ and variation $= \theta$.

$\lambda_i$ is the individual's event rate, described as rate ratios from the group average rate in the extended Poisson model. Right panel gives an example of the Gamma distribution, which is used to model the distribution of the frailty variable (mean one and variance $\theta$) in the negative binomial distribution.

The individual’s event rate ($Z \cdot \lambda$) is an average of the event rates in the group ($\lambda$) multiplied by the relative size of the individual rate compared with the estimated average event rate in the group ($Z$). The variable $T$ denotes the individuals time at risk.

Fitting the negative binomial distribution to hospital admission data (Table 3, first row) yields an apparently good visual agreement between observed data and expected values from the negative binomial distribution (Fig. 3), but the difference between the observed and the fitted values reached statistical significance ($P < 0.001$) due to the large number of data. However, frailty models like the negative binomial regression model seem to give a good description of our data. Other studies have also found that the negative binomial distribution provides a good description of both hospital admissions and consultations in general practice.

<table>
<thead>
<tr>
<th>Multiple events</th>
<th>Group</th>
<th>Admission rates per person-years</th>
<th>Rate ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative binomial regression</td>
<td>Control</td>
<td>0.166 (0.116–0.237)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Intervention 1</td>
<td>0.136 (0.103–0.179)</td>
<td>0.81 (0.52–1.29)</td>
</tr>
<tr>
<td></td>
<td>Intervention 2</td>
<td>0.138 (0.112–0.169)</td>
<td>0.83 (0.55–1.25)</td>
</tr>
<tr>
<td>Poisson regression with random effect</td>
<td>Control</td>
<td>0.167 (0.125–0.222)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Intervention 1</td>
<td>0.141 (0.106–0.189)</td>
<td>0.85 (0.56–1.27)</td>
</tr>
<tr>
<td></td>
<td>Intervention 2</td>
<td>0.141 (0.106–0.189)</td>
<td>0.85 (0.56–1.28)</td>
</tr>
<tr>
<td>Standard Poisson regression with robust variance estimation</td>
<td>Control</td>
<td>0.163 (0.114–0.233)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Intervention 1</td>
<td>0.128 (0.100–0.163)</td>
<td>0.78 (0.51–1.21)</td>
</tr>
<tr>
<td></td>
<td>Intervention 2</td>
<td>0.131 (0.109–0.158)</td>
<td>0.81 (0.54–1.20)</td>
</tr>
</tbody>
</table>

The negative binomial distribution model is appealing as it naturally accommodates the different probabilities for events across members of a population. Moreover, the distribution of the variation of individual rate ratios is depicted as a gamma distribution, which is mathematically convenient and a highly
flexible distribution. However, inter-individual frailty variations could, indeed, follow other distributions. It may, however, require a large sample to distinguish between the negative binomial model and a Poisson model with a different frailty distribution.

There may be other sources of additional variability than from the recurrent events among individuals. The individuals are clustered within GPs, and GPs may be another level of extra variability. In the present study we examined this by assigning a level to each practice and found no extra variability among practices. There was not data available to allow stratification within GPs. In other settings correlation of events within practices or GPs may result in additional variation that needs to be accounted for in the analysis.

Other approaches to the analysis of recurrent events
Assumptions of the underlying distribution of the variation of the individual rate ratios may be avoided using a pseudo-likelihood method by Carroll and Ruppert.15 i.e.

\[
\text{Po}(Z \cdot \lambda \cdot T), \text{where no assumption of Z is made.}
\]

This approach solves the challenges described earlier, i.e. it handles incomplete follow-up of some individuals and dependence between recurrent events using an individual frailty for which a special distribution is not required (Table 3, second row).

An alternative to the frailty models described above is to use the standard Poisson model with robust variance estimation. As previously discussed, the first event with its corresponding time at risk could be analysed using a standard Poisson regression model. So could the second event with its corresponding time at risk. Individuals who experience no (first) event would not contribute in the analysis of the second event. The analysis of first event, second event and so forth are, of course, correlated since they use data on the same individuals. This correlation can be taken into account using robust variance estimation, which estimates the correlation between observations on the same individuals and corrects for the correlation. It is, however, important to note that there is a selection of individuals with time; all individuals contribute in the analysis of individuals with time; all individuals contribute in the first analysis, but only individuals experiencing at least one event contribute in the second analysis and so forth. Such selection may create bias,16 and some care should therefore be exercised when analysing recurrence with robust variance estimation: In the hospital admission data, the selection seemed to be balanced in the three randomized groups, and we obtained comparable parameter estimates (last row in Table 3).

Conclusion
Calculation of rate ratios under the assumption that data follow a specific mathematical distribution generally enhances the statistical precision of the estimate. It is, however, important that the theoretical model fits data adequately and affords straightforward and intuitive data interpretation. An individual rate model that includes a parameter of an unspecified individual event distribution frailty may be a natural choice when analysing longitudinal data of contacts to the health care system in broad terms.

Evaluation of health care contacts from first events alone often misses large amounts of potentially important data and may produce different results than evaluation including all (recurrent) events. Analysis of health care contacts should therefore embrace both first and recurrent events and it should use a model appropriate to these data.

Declaration
Funding: ETP is employed at the Institute of Public Health, Aarhus University; JLT is funded by a grant from the Danish National Research Foundation. Ethical approval: permission to conduct register investigation in relation to the Ebeltoft Health Promotion Project was given by the Scientific Ethical Committee of Aarhus County (J. no. 1990/1966) and the Danish Data Protection Agency (J. no. 2001-41-0738).

Conflict of interest: none.

Acknowledgements
We thank the Ebeltoft Health Promotion Project for providing data for the examples in this study.

References


