Heterogeneity and event dependence in the analysis of sickness absence

Isabel Torá-Rocamora1,2, David Gimeno1,2,3, George Delclos1,2,4, Fernando G Benavides1,2, Rafael Manzanera5, Josefina Jardí5, Constança Alberti5, Yutaka Yasui6† and José Miguel Martínez1,2*†

Abstract

Background: Sickness absence (SA) is an important social, economic and public health issue. Identifying and understanding the determinants, whether biological, regulatory or, health services-related, of variability in SA duration is essential for better management of SA. The conditional frailty model (CFM) is useful when repeated SA events occur within the same individual, as it allows simultaneous analysis of event dependence and heterogeneity due to unknown, unmeasured, or unmeasurable factors. However, its use may encounter computational limitations when applied to very large data sets, as may frequently occur in the analysis of SA duration.

Methods: To overcome the computational issue, we propose a Poisson-based conditional frailty model (CFPM) for repeated SA events that accounts for both event dependence and heterogeneity. To demonstrate the usefulness of the model proposed in the SA duration context, we used data from all non-work-related SA episodes that occurred in Catalonia (Spain) in 2007, initiated by either a diagnosis of neoplasm or mental and behavioral disorders.

Results: As expected, the CFPM results were very similar to those of the CFM for both diagnosis groups. The CPU time for the CFPM was substantially shorter than the CFM.

Conclusions: The CFPM is an suitable alternative to the CFM in survival analysis with recurrent events, especially with large databases.

Keywords: Sickness absence, Survival analysis, Conditional frailty model, Poisson regression, Mental disorders, Neoplasms

Background

Sickness absence (SA) is a complex phenomenon with great economic and social impact, and is considered a major occupational and public health issue [1-3]. SA is defined as a temporary situation in which a worker is unable to perform his/her usual work, either because of illness or injury [4]. The duration of SA affects the individual worker’s quality of life, and have a great impact in his/her family, employer and society overall [5]. Knowing what factors are associated with how long a sickness absence episode lasts is of great importance in trying to reduce the SA duration. Sickness absence duration has been examined using a number of statistical techniques, most frequently survival analysis techniques [6-8]. Generally, survival studies analyze the time until the occurrence of a certain event of interest (e.g., death) [9]. However, in the context of sickness absence, some individuals may be more prone to experience multiple events, whether due to new illnesses or injuries, or recurrence of the same event. Repeated events can create within-subject correlation in event times [8,10-12], arising from two sources: 1) event dependence; and 2) heterogeneity across individuals [11]. Event dependence occurs when the risk of a particular event depends on events previously experienced, whereas heterogeneity occurs when some individuals have a higher or lower risk of experiencing the events due to unknown, unmeasured or unmeasurable factors. Consequently, analytical approaches to...
modeling of sickness absence duration should take into account both event dependence and heterogeneity to avoid obtaining biased estimates of the parameters of interest [11,12].

The conditional frailty model (CFM) proposed by Box-Steffensmeier and De Boef [11], which can be viewed as an extension of the Cox model, simultaneously captures event dependence and heterogeneity [11], and has been used previously in political sciences research [12]. The computational applicability of the CFM maybe limited when dealing with very large datasets such a sickness absence registries, numbering hundreds of thousands or millions of individuals and/or episodes. For example, in Catalonia for the year 2007, the Catalan Institute of Medical and Health Evaluations (ICAMS, by its Spanish acronym) recorded 800,464 sickness episodes in 580,959 persons. It is well established that Poisson regression is a possible alternative to Cox regression [13,14]. Specifically, when a Cox model is confronted with computational limitations in analyzing large databases, a Poisson regression model maybe a reasonable alternative [15].

The goal of this paper is to propose a Poisson-based conditional frailty model (CFPM) that accounts for event dependence and heterogeneity for a large data analysis of sickness absence.

**Methods**

In the first section we will introduce the CFM and explain the proposed CFPM. In the following section we will explain the methods used to empirically compare the CFM and CFPM.

**Conditional frailty and conditional frailty Poisson models**

**The conditional frailty model**

The CFM models the dependence of events and heterogeneity by stratifying the baseline hazard function by event order and incorporating random effects for individuals, respectively. The formulation of the model is in gap time data structure incorporated in (t−tk−1). Let the hazard of k-th event occurring in the i-th individual, the CFM is defined as

\[ \lambda_{ik}(t) = \lambda_{0k}(t-t_{k-1})e^{X_i\beta+\omega_i} \]

where tk−1 is the time of occurrence of (k−1)th event, \( \lambda_{0k} \) is the baseline hazard rate of the k-th event, \( \beta \) is the vector of parameters associated with covariates \( X \) and \( \omega_i \) is the random effect or “frailty” of the i-th individual that follows a gamma distribution. Considering right-censored failures, the parameters are interpreted as the log hazard ratio estimates associated with covariates for an event since the previous event, due to the gap time data structure incorporated in (t−tk−1). More details about the CFM can be found in Box-Steffensmeier et al. [11,12].

**The conditional frailty Poisson model**

The CFPM considers \( \lambda_{ik}(t) \) to be the hazard of k-th event at time t occurring in the i-th individual, as

\[ \lambda_{ik}(t) = \lambda_{0k}(t-t_{k-1})e^{X_i\beta+\omega_i} \]

Following the piecewise exponential model formulation [16], the baseline hazard for the k-th event is defined as

\[ \lambda_{0k}(t) = \sum_{j=1}^{J} \lambda_{0k}^* I\{t_{j-1}, t_j]\]

with divisions of time scale into \( (t_1, t_2), (t_2, t_3), \ldots, (t_{j-1}, t_j) \) which are J non-zero, nonoverlapping intervals, with \( t_1 = 0 \). The model captures event dependence (i.e., the dependence of the risk of a subsequent event on the occurrences of previous events) by allowing the baseline hazard to vary by event orders using an index “k” for the baseline hazard \( \lambda_{0k}^* \) for the k-th event. The heterogeneity is controlled by including an \( \omega_i \) random effect for the i-th individual. We consider a gamma distribution for the random effect.

Let \( n_{ijk} \) and \( d_{ijk} \) denote the time at risk and a covariate indicator of an event \( (d_{ijk} = 1) \) or non-event \( (d_{ijk} = 0) \), in the j-th time interval, for i-th individual and k-th event. The proposed Poisson regression model assumes a Poisson distribution on \( d_{ijk} \) with the following log-linear mean,

\[ \log(E[d_{ijk} | \omega_i]) = \log(\lambda_{ik}(t)) + \log(n_{ijk}) \]

Note that the observed duration of SA (“time at risk under observation”) is include as an offset term in the Poisson model which starts on the day of SA certification and ends on the day the worker returns to work or the day the worker’s SA status becomes unknown (e.g., due to retirement, death, emigration), whichever is earlier.

**Empirical comparison between conditional frailty models**

**Description of the data**

The CFM and CFPM were compared empirically using data from all episodes of non work-related SA that occurred in Catalonia (Spain) in 2007 (n = 800,464). Specifically, we assessed the influence of certain covariates of interest on SA duration, where the end of the episode of SA is considered the event of interest. A same individual may suffer more than one SA during the study period and therefore SAs are repeatable events.

The data were recorded through the Integrated Management System for Sickness Absence (SIGIT, by its Spanish acronym) at the ICAMS, a computerized registry...
and connected to all physicians in Catalonia responsible for certifying SA episodes.

For each episode, the diagnosis at case closure was available, coded according to the International Classification of Diseases, 10th Edition (ICD-10). We separately analyzed two large ICD-10 diagnosis groups selected to reflect frequent SA diagnoses (mental and behavioral disorders, codes F00-F99) and SA diagnoses with typically long duration times (neoplasms, codes C00-D48). Mental and behavioral disorders accounted for 3,268,075 days from 59,647 episodes in 53,238 individuals with a median duration of 10 days (25th percentile, 25 days; 75th percentile, 67 days); and neoplasms accounted for 516,676 days from 7,431 episodes in 6,975 individuals with median duration of 11 days (25th percentile, 28 days; 75th percentile, 80 days). Approximately 10% of individuals had repeated events. For neoplasms, repeated events occur in 5% of individuals. Problems with convergence may emerge if there are too many event-order strata and/or a small number of episodes per stratum in both CFM [12] and CFPM. Therefore, we collapsed the event number so that any number of repeated episodes greater than 5 was set equal to 5.

Other covariates of interest were sex, age (16–28, 29–35, 36–45, >45 years), economic activity (11 branches), Catalan health region, entity responsible for case management (National Institute of Social Security or a mutual insurance company), and employment status (salaried or self-employed).

**Empirical comparison**

We empirically compared the hazard ratio (HR) and 95% confidence intervals (95% CI) obtained by the CFM and the proposed CFPM. To define the baseline hazard function in the CFPM following the piecewise exponential model, we chopped time into 90-day-length non-overlapping.

To explore the source of correlation existing in the data and to better assess the proposed CFPM as a reliable alternative to the CFM, we also computed the HR and 95% CI, with models which: 1) only take into account the event dependence; or 2) only take into account for heterogeneity. The former models were based on a gap time conditional model (CM) [17] which takes into account the event dependence by stratifying the baseline hazard function according to event order [18]. The CM is similar to CFM but does not include the individual random effect term. We also ran a conditional Poisson model (CPM) with the same expression as the CFPM, but without the random effect term by individual. With respect to models that control only for heterogeneity we considered a frailty model (FM), which is similar to the CFM but without stratifying the baseline hazard functions by event order and controls for the heterogeneity by including random effects for individuals.

Finally, we ran a Poisson model that takes into account only heterogeneity (FPM). The FPM presents a similar expression to the CFPM, but without the interaction between event order and the baseline hazard function.

Based on Box-Steffensmeier and De Boef [11] we hypothesized that when event dependence is strong, the event-dependence-only models (CM and CPM) should give estimates of the effects which are closed to the CFM, than models that do not control for the dependence of events (FM and FPM). Similarly, if heterogeneity is strong, the results of frailty models (FM and FPM) should be closer to the CFM than the models which only take into account dependence of events (CM and CPM). For both cases, i.e., regardless of the cause of correlation that predominates (event dependence or heterogeneity), we should expect that the estimates of CFPM will be closer to the CFM than the other models that only control for event dependence or only for heterogeneity. Thus, the comparison of the different models with the CFM serves to evaluate the suitability of CFPM when there is event dependence and/or heterogeneity.

The results between models were compared using the % relative bias (%RB) in point estimate and the % relative width difference in confidence interval (%RW), using the CFM as reference [15]. These measures are defined as

\[
\% \text{RB} = \left( \frac{HR_{\text{Other}} - HR_{\text{CFM}}}{HR_{\text{CFM}}} \right) \times 100
\]

\[
\% \text{RW} = \left( \frac{(U_{\text{Other}} - L_{\text{Other}}) - (U_{\text{CFM}} - L_{\text{CFM}})}{(U_{\text{CFM}} - L_{\text{CFM}})} \right) \times 100
\]

where \(HR_{\text{Other}}\) and \(HR_{\text{CFM}}\) are the hazard ratio under a specific model (CM, FM, CPM, FPM, CFPM) and the CFM, respectively, and \(U_{\text{Other}}\) and \(L_{\text{Other}}\) are the respective upper and lower confidence interval endpoints a specific model, and \(U_{\text{CFM}}\) and \(L_{\text{CFM}}\) represent the upper and lower confidence interval endpoints for the CFM, respectively.

To compare the time for obtaining the parameter estimates from CFPM and CFM, their respective CPU time was measured on the Windows XP operating system on Intel® Core™2 CPU machine. The CFM and CFPM were fitted using R version 2.8.1 and Stata version 11, respectively. The Stata code for the CFPM, FPM and CPM is provided in the Additional file 1. Specifically we used the poisson command for the CPM, and the xtpoisson command for the CFPM and the FPM. Information about these commands can be found in the book written by Rabe-Hesketh S and Skrondal A [19].

This study was approved by the Parc de Salut Mar Clinical Research Ethical Committee of Barcelona, Spain (number 2011/4229/I).
Results
Tables 1 and 2 summarize the HR estimates and 95% CIs, for the six models, adjusted for covariates, for mental or behavioural disorders and neoplasms, respectively. The six models we considered showed associations that were in the same direction (for a specific group of a covariate the HR were above 1 (or below 1) in the six models). Although the associations for all six models were in the same direction, there were differences in the magnitude of HR across the models. The CFPM results were very similar to those of the CFP for both diagnosis groups (Tables 3 and 4).

For neoplasms, the %RB for the CFPM ranged from 0% to 6.9% (absolute values), and the %RW from 0% to 6.9% (Tables 3 and 4).

### Table 1 Hazard ratio and 95% confidence interval for selected covariates from episodes of non work-related sickness absence occurred in Catalonia (Spain) in 2007 for mental or behavioural disorders (n = 59,647)

| Health region                  | CFPM | CPM | FPM | CM | FM | CFM |
|-------------------------------|------|-----|-----|----|----|-----|
| Barcelona                     | 1.00 (−) | 1.00 (−) | 1.00 (−) | 1.00 (−) | 1.00 (−) | 1.00 (−) |
| Lleida                        | 1.05 (0.98-1.12) | 1.05 (1.00-1.11) | 1.04 (0.97-1.12) | 1.04 (1.00-1.09) | 1.04 (0.97-1.12) | 1.05 (0.98-1.11) |
| Camp de Tarragona              | 1.09 (1.04-1.15) | 1.03 (0.98-1.07) | 1.10 (1.05-1.16) | 1.04 (1.00-1.08) | 1.16 (1.09-1.23) | 1.07 (1.02-1.12) |
| Terres de l’Ebre               | 1.09 (0.99-1.21) | 1.06 (0.98-1.15) | 1.09 (0.99-1.20) | 1.05 (0.97-1.13) | 1.10 (0.98-1.23) | 1.08 (0.98-1.18) |
| Girona                        | 1.01 (0.97-1.05) | 1.01 (0.98-1.04) | 1.01 (0.97-1.05) | 1.01 (0.98-1.04) | 1.00 (0.96-1.05) | 1.01 (0.97-1.04) |
| Catalunya Central             | 1.06 (1.01-1.11) | 1.04 (1.00-1.08) | 1.06 (1.01-1.11) | 1.04 (1.00-1.07) | 1.06 (1.01-1.12) | 1.05 (1.01-1.10) |
| Alt Pirineu i Aran            | 0.99 (0.79-1.25) | 0.95 (0.78-1.16) | 0.99 (0.78-1.25) | 0.94 (0.79-1.13) | 1.01 (0.78-1.31) | 0.97 (0.79-1.19) |
| Economic activity branch      |      |      |      |    |    |      |
| Agriculture, mining, fishing  | 1.04 (0.71-1.52) | 0.99 (0.72-1.35) | 1.05 (0.72-1.55) | 1.01 (0.76-1.32) | 1.09 (0.71-1.67) | 1.03 (0.73-1.45) |
| Manufacturing industry, energy production | 1.02 (0.70-1.49) | 0.99 (0.72-1.36) | 1.02 (0.69-1.50) | 1.01 (0.76-1.33) | 1.05 (0.68-1.51) | 1.01 (0.72-1.43) |
| Construction                  | 0.89 (0.61-1.30) | 0.88 (0.64-1.20) | 0.91 (0.69-1.19) | 0.91 (0.69-1.39) | 0.89 (0.64-1.26) | 0.91 (0.69-1.39) |
| Commercial/vehicles repair    | 0.88 (0.60-1.29) | 0.87 (0.63-1.19) | 0.90 (0.68-1.18) | 0.90 (0.58-1.38) | 0.89 (0.63-1.25) | 0.90 (0.58-1.38) |
| Hotel, restaurant businesses  | 0.91 (0.62-1.34) | 0.90 (0.66-1.24) | 0.93 (0.63-1.36) | 0.93 (0.70-1.22) | 0.94 (0.61-1.44) | 0.92 (0.65-1.30) |
| Transportation/communication  | 0.92 (0.63-1.35) | 0.91 (0.66-1.24) | 0.93 (0.63-1.37) | 0.93 (0.71-1.23) | 0.95 (0.62-1.46) | 0.93 (0.66-1.30) |
| Finance, real estate, services| 0.90 (0.62-1.31) | 0.89 (0.65-1.22) | 0.91 (0.62-1.34) | 0.92 (0.70-1.21) | 0.92 (0.60-1.42) | 0.91 (0.64-1.27) |
| Government                    | 0.95 (0.65-1.39) | 0.95 (0.69-1.30) | 0.96 (0.65-1.41) | 0.97 (0.73-1.27) | 0.98 (0.64-1.50) | 0.96 (0.68-1.35) |
| Health, education, other social activities | 0.92 (0.53-1.60) | 0.94 (0.61-1.44) | 0.91 (0.52-1.59) | 0.96 (0.56-1.41) | 0.93 (0.50-1.74) | 0.94 (0.57-1.54) |
| Domestic housekeeping         | 0.69 (0.39-1.22) | 0.72 (0.49-1.08) | 0.69 (0.39-1.23) | 0.77 (0.54-1.10) | 0.68 (0.36-1.30) | 0.73 (0.44-1.22) |
| Extraterritorial agencies     |      |      |      |    |    |      |

### Abbreviations:
- HR: hazard ratio
- 95% CI: 95% confidence interval
- CFPM: conditional frailty Poisson model
- CPM: conditional poisson model
- FPM: frailty poisson model
- CM: conditional model
- FM: frailty model
- CFM: conditional frailty model

Page 4 of 9
http://www.biomedcentral.com/1471-2288/13/114
Torá-Rocamora et al. BMC Medical Research Methodology 2013, 13:114
### Table 2 Hazard ratio and 95% confidence interval for selected covariates from episodes of non work-related sickness absence occurred in Catalonia (Spain) in 2007 for neoplasms (n = 7,431)

| Health region | Gender | Age at onset of absence episode (years) | Economic activity branch | Social Security regime | Entity manage |
|---------------|--------|----------------------------------------|--------------------------|------------------------|---------------|
|               |        | 16 – 28                                | 29 – 35                  | 36 – 45                | > 45          |
|               |        | 1 – 15                                  | 16 – 45                  | 16 – 45                | 1 – 15        |
|               |        | Male                                   | Female                   | Male                   | Female       |
|               |        | 1.00 (–)                               | 0.92 (0.85-0.99)         | 1.00 (–)               | 0.92 (0.85-0.99) |
|               |        | 1.00 (–)                               | 0.95 (0.89-1.01)         | 1.00 (–)               | 0.95 (0.89-1.01) |
|               |        | 1.00 (–)                               | 0.92 (0.86-0.99)         | 1.00 (–)               | 0.92 (0.86-0.99) |
|               |        | 1.00 (–)                               | 0.95 (0.90-1.00)         | 1.00 (–)               | 0.95 (0.90-1.00) |
|               |        | 0.98 (0.80-1.00)                       | 0.91 (0.90-1.00)         | 0.91 (0.84-0.98)       | 0.91 (0.84-0.98) |
| 1 – 15         | Female | 1.00 (–)                               | 0.93 (0.86-0.99)         | 1.00 (–)               | 0.93 (0.86-0.99) |
|               |        | 1.00 (–)                               | 0.96 (0.89-1.01)         | 1.00 (–)               | 0.96 (0.89-1.01) |
|               |        | 1.00 (–)                               | 0.97 (0.90-1.00)         | 1.00 (–)               | 0.97 (0.90-1.00) |
|               |        | 1.00 (–)                               | 0.96 (0.90-1.00)         | 1.00 (–)               | 0.96 (0.90-1.00) |
|               |        | 0.98 (0.90-1.00)                       | 0.91 (0.84-0.98)         | 0.91 (0.84-0.98)       | 0.91 (0.84-0.98) |
|               |        | 1.00 (–)                               | 0.94 (0.88-0.99)         | 1.00 (–)               | 0.94 (0.88-0.99) |
|               |        | 1.00 (–)                               | 0.96 (0.90-1.00)         | 1.00 (–)               | 0.96 (0.90-1.00) |
|               |        | 1.00 (–)                               | 0.97 (0.90-1.00)         | 1.00 (–)               | 0.97 (0.90-1.00) |
|               |        | 0.98 (0.90-1.00)                       | 0.91 (0.84-0.98)         | 0.91 (0.84-0.98)       | 0.91 (0.84-0.98) |

**Abbreviations:** HR hazard ratio, 95 CI% confidence intervals at 95%, CFPM conditional frailty poisson model, CPM conditional poisson model, FPM frailty poisson model, FM frailty model, CFM conditional frailty model.

7.1%. For the FPM and FM, these results were not as close to the CFM as the CFPM (10.3% in the %RB and 14.3% in the %RW for age > 45 in the FPM, and ~17.9% in the %RB for extraterritorial agencies, and 33.3% in the %RW for Camp de Tarragona Health Region in the FM). The results for the CM and CPM were further apart from the CFM as compared to the FPM and FM, in some cases %RB reaching the 20%, 40% or, in the case of CM, 60% and %RW exceeding the 20%.

For mental or behavioural disorders, the CFPM, CPM and FPM behaved similar and were better than the CM and FM, and the CFPM behaved very closely to the
In terms of %RW in general the CFPM presents the lowest percentages, but they can be up to 15%. In the case of CM and FM, the %RW can reach 30-40%.

The CPU time for the CFPM was much shorter than the CFM. Using R version 2.8.1. on the Windows XP operating system on Intel® Core™2 CPU machine, the CFM took 124,877.67 (2,081.30 minutes) and 647.53 seconds (10.80 minutes) CPU time for mental health disorders and neoplasm data analysis, respectively. Using Stata version 11, on the same operating system and

| Gender         | CFPM | CPM | FPM | CM | FM |
|----------------|------|-----|-----|----|----|
| Male           | -    | -   | -   | -  | -  |
| Female         | -1.1 | 25.0| 1.1 | 0.0| -2.2| 0.0| 1.1 | -25.0| -3.3| 25.0|

### Age at onset of absence episode (years)

| Age at onset of absence episode (years) | CFPM | CPM | FPM | CM | FM |
|----------------------------------------|------|-----|-----|----|----|
| 16 - 28                                | -    | -   | -   | -  | -  |
| 29 - 35                                | -1.2 | 0.0 | 1.2 | -16.7 | 2.8 | 10.0 | -2.8 | -20.0 | 8.4 | 40.0 |
| 36 - 45                                | -2.6 | 0.0 | 3.9 | 0.0 | -2.6 | 0.0 | 5.1 | -20.0 | -6.4 | 0.0 |
| > 45                                   | -4.6 | 0.0 | 6.1 | 0.0 | -4.6 | 25.0 | 9.1 | 0.0 | -10.6 | 25.0 |

### Health region

| Health region   | CFPM | CPM | FPM | CM | FM |
|-----------------|------|-----|-----|----|----|
| Barcelona       | -    | -   | -   | -  | -  |
| Lleida          | 0.0  | 7.7 | 0.0 | -15.4 | -1.0 | 15.4 | -1.0 | -30.8 | -1.0 | 15.4 |
| Camp de Tarragona | 1.9  | 10.0 | -3.7 | -10.0 | 2.8 | 10.0 | -2.8 | -20.0 | 8.4 | 40.0 |
| Terres de l'Ebre | 0.9  | 10.0 | -1.9 | -15.0 | 0.9 | 5.0 | -2.8 | -20.0 | 1.9 | 25.0 |
| Girona          | 0.0  | 14.3 | 0.0 | -14.3 | 0.0 | 14.3 | 0.0 | -14.3 | -1.0 | 28.6 |
| Catalunya       | 1.0  | 11.1 | -1.0 | -11.1 | 1.0 | 11.1 | -1.0 | -22.2 | 1.0 | 22.2 |
| Alt Pirineu     | 2.1  | 15.0 | -2.1 | -5.0 | 2.1 | 17.5 | -3.1 | -15.0 | 4.1 | 32.5 |

### Economic activity branch

| Economic activity branch                        | CFPM | CPM | FPM | CM | FM |
|-----------------------------------------------|------|-----|-----|----|----|
| Agriculture, mining, fishing                  | -    | -   | -   | -  | -  |
| Manufacturing industry, energy production     | 1.0  | 12.5 | -3.9 | -12.5 | 1.9 | 15.3 | -1.9 | -22.2 | 5.8 | 33.3 |
| Construction                                  | 1.0  | 11.3 | -2.0 | -9.9 | 1.0 | 14.1 | 0.0 | -19.7 | 4.0 | 31.0 |
| Commercial/vehicles repair                    | 0.0  | 11.3 | -1.1 | -9.7 | 0.0 | 12.9 | 2.3 | -19.4 | 2.3 | 29.0 |
| Hotel, restaurant businesses                  | -1.1 | 11.3 | -2.3 | -9.7 | -1.1 | 12.9 | 1.1 | -19.4 | 1.1 | 29.0 |
| Transportation/communication                 | -1.1 | 10.8 | -2.2 | -10.8 | 0.0 | 12.3 | 1.1 | -20.0 | 2.2 | 27.7 |
| Finance, real estate, services                | -1.1 | 12.5 | -2.2 | -9.4 | 0.0 | 15.6 | 0.0 | -18.8 | 2.2 | 31.3 |
| Government                                   | -1.1 | 9.5  | -2.2 | -9.5 | 0.0 | 14.3 | 1.1 | -19.1 | 1.1 | 30.2 |
| Health, education, other social activities    | -1.0 | 10.5 | -1.0 | -9.0 | 0.0 | 13.4 | 1.0 | -19.4 | 2.1 | 28.4 |
| Domestic housekeeping                         | -2.1 | 10.3 | 0.0 | -14.4 | -3.2 | 10.3 | 2.1 | -22.7 | -1.1 | 27.8 |
| Extraterritorial agencies                    | -5.5 | 6.4  | -1.4 | -24.4 | -5.5 | 7.7  | 5.5 | -28.2 | -6.9 | 20.5 |

### Social Security regime

| Social Security regime                        | CFPM | CPM | FPM | CM | FM |
|-----------------------------------------------|------|-----|-----|----|----|
| Under contract                                | -    | -   | -   | -  | -  |
| Self-employed                                 | -8.5 | 0.0 | 11.3 | -25.0 | -9.9 | 0.0 | 12.7 | -37.5 | -18.3 | 0.0 |

### Entity manage

| Entity manage                                  | CFPM | CPM | FPM | CM | FM |
|-----------------------------------------------|------|-----|-----|----|----|
| National Institute of Social Security         | -    | -   | -   | -  | -  |
| Insurance company                             | 1.8  | 0.0 | -0.9 | 0.0 | 1.8 | 20.0 | -2.7 | -40.0 | 2.7 | 40.0 |

**Abbreviations:** %RB % relative bias, %RW % relative width, CFPM conditional frailty poisson model, CPM conditional poisson model, FPM frailty poisson model, CM conditional model, FM frailty model, CFM conditional frailty model.
The CFPM took 260.56 (4.34 minutes) and 35.77 (0.60 minutes) seconds for mental health disorders and neoplasm, respectively.

**Discussion**

We proposed for the first time a Poisson-based conditional frailty model that accounts for both event dependence and heterogeneity. The CPU time required for the CFPM was substantially shorter than that for the CFM. In addition, as expected, the CFPM results were very similar to those of the CFM for both diagnosis groups.

The similarity of results between the CFM and CFPM, and the differences noted with models that do not include event dependence and/or heterogeneity reinforces...
the usefulness of the CFPM. In the case of neoplasms, the %RB for frailty models is closer to the CFM than for conditional models, suggesting that the dependence that dominates the data is heterogeneity. Conversely, in the case of mental health disorders, the %RB is smaller in the CM than that of the FM, indicating a greater influence of event dependence.

The choice of time intervals may influence the model fit result. The key issue is to sufficiently approximate the underlying hazard function over time by a set of piecewise-constant hazards in Poisson models. The shorter we make the time intervals of the piecewise-constant hazards the closer Poisson models get to Cox models. If data in each time interval become sparse by making the intervals shorter, however, parameter estimation becomes unstable, which in turn affect the estimation of the covariates’ effects of interest. As Michael Friedman pointed out “precise practical guidelines for choosing the number of intervals have not been formulated” [20]. Choosing different cut-points has a trade-off. It will be helpful to explore the form of the underlying hazard function and also assess the availability of data in each interval. In addition, performing a sensitivity analysis choosing different cut-points is useful for assessing changes in the parameters estimates of interest.

To avoid convergence problems we treated repeated episodes greater than 5 as equal to 5. The percentage of individuals with more than 5 repeated episodes for neoplasms is 0.52%, and for mental and behavioural disorders is 0.35%. Due to the very low percentages of individuals with more than 5 episodes, treating episodes greater than 5 as equal to 5 do not change the results.

A key advantage of the CFPM over the CFM is the reduction of computational time when analyzing large databases. This may be particularly important for institutions in countries where computers with high computational speed are not readily available. Currently, the CFM can only be run using R version 2.8.1. software [21]. The CFPM, though, can easily be run using other, statistical software such as Stata [22].

Conclusions
In summary, assuming that within-subject correlation is a result of event dependence will result in biased estimates when, in fact, it is due to heterogeneity in the data. Conversely, assuming correlation in event times is due to heterogeneity will also result in biased inferences when, in fact, the source is event dependence [12]. For this reason, we recommend incorporating both sources of correlation when fitting a model. To achieve this, the CFPM is an attractive alternative to the CFM in survival analysis with recurrent events, especially with large databases, such as those that may exist for the analysis of sickness absence data.

Additional file

Additional file 1: The file includes the Stata syntax for Conditional Frailty Poisson Model (CFPM), Frailty Poisson Model (FPM) and Conditional Poisson Model (CPM).

Abbreviations
SA: Sickness absence; ICAMS (by its Spanish acronym): Catalan Institute of Medical and Health Evaluations; CFM: Conditional frailty model; CFPM: Poisson-based conditional frailty model; SIGIT (by its Spanish acronym): Integrated management system for sickness absence; ICD-10: International classification of diseases 10th edition; HR: Hazard ratio; CM: Conditional model; CPM: Conditional poisson model; FM: Frailty model; FPM: Frailty poisson model; %RB: % Relative bias; %RW: % Relative width difference in confidence interval; CPU: Central processing unit.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
ITR, YY, and JMM conceived the idea and worked on in the statistical and public health issues of the method, conducted the data analysis, and wrote the initial draft. DG, GD, and FGB gave critiques from public-health points of view and study design. JJ, RM, and CA provided discussion based on sickness absence. CA assisted the data analyses. All authors reviewed numerous drafts of the manuscript, are in agreement with the text and findings, and we have all approved this final version.

Acknowledgements
This work was supported by grants from the Fondo de Investigación Sanitaria (PI11/01470), Operating Grant of the Canadian Institute of Health Research entitled ‘Statistical Methods for Epidemiologic Investigations’ and the Institut Català d’Avaluacions Mèdiques. The authors would like to thanks the Institut Català d’Avaluacions Mèdiques i Sanitàries (ICAMS) for providing the database.

Author details
1Department of Experimental and Health Sciences, Center for Research in Occupational Health (CISAL), Universitat Pompeu Fabra (UPF), C/Doctor Aiguader 88, Barcelona 08003, Spain. 2CIBER in Epidemiology and Public Health, Barcelona, Spain. 3Southwest Center for Occupational and Environmental Health, Division of Epidemiology, Human Genetics and Environmental Sciences, The University of Texas School of Public Health, San Antonio Campus, San Antonio, TX, USA. 4Southwest Center for Occupational and Environmental Health, Division of Epidemiology, Human Genetics and Environmental Sciences, The University of Texas School of Public Health, Houston Campus, Houston, TX, USA. 5Institut Català d’Avaluacions Mèdiques i Sanitàries (ICAMS), Departament de Salut, Generalitat de Catalunya, Barcelona, Spain. 6Department of Public Health Sciences, School of Public Health, University of Alberta, Edmonton, AB, Canada.

Received: 18 April 2013 Accepted: 11 September 2013
Published: 16 September 2013

References
1. Moncada S, Navarro A, Coríes I, Molinero E, Artazaos L: Sickness leave, administrative category and gender: results from the “Casa Gran” project. Scand J Public Health Suppl 2002, 59:26–33.
2. Whitaker SC: The management of sickness absence. Occup Environ Med 2001, 58:420–424.
3. Gimeno D, Benavides FG, Benach J, Amick BC 3rd: Distribution of sickness absence in the European union countries. Occup Environ Med 2004, 61:867–869.
4. Fábrega O, Company A: La gestión de la incapacidad temporal en Gestión de Atención Primaria. In Curso a distancia de Gestión de Atención Primaria. Edited by Gené Badia J, Grego Recasens JM. Barcelona: semFYC y Formación Continuada Les Hueres – Universitat de Barcelona. Fundació Bosch Gimpera; 1999:95–103.
5. Alexanderson K: Sickness absence: a review of performed studies with focused on levels of exposures and theories utilized. Scand J Soc Med 1998, 26:241–249.
6. Navarro A, J Reis R, Martin M: Some alternatives in the statistical analysis of sickness absence. Am J Ind Med 2009, 52:811–816.
7. González JR, Peña EA: Non-parametric estimation of survival function for recurrent events data. Rev Esp Salud Publica 2004, 78:189–199.
8. Roelen CA, Koopmans PC, Anema JR, van der Beek AJ: Recurrence of medically certified sickness absence according to diagnosis: a sickness absence register study. J Occup Rehabil 2010, 20:113–121.
9. Barceló MA: Marginal and conditional models in multivariate survival analysis. Gac Sanit 2002, 16:59–68.
10. Christensen KB, Andersen PK, Smith-Hansen L, Nielsen ML, Kristensen TS: Analyzing sickness absence with statistical models for survival data. Scand J Work Environ Health 2007, 33:233–239.
11. Box-Steffensmeier JM, De Boef S: Repeated events survival models: the conditional frailty model. Stat Med 2006, 25:3518–3533.
12. Box-Steffensmeier JM, De Boef S, Joyce KA: Event dependence and heterogeneity in duration models: the conditional frailty model. Polit Anal 2007, 15:237–256.
13. Aitkin M, Clayton D: The fitting of exponential, weibull and extreme value distributions to complex censored survival data using GLIM. Appl Statist 1980, 29:156–163.
14. Laird N, Olivier D: Covariance analysis of censored survival data using log-linear analysis techniques. J Am Stat Assoc 1981, 76:231–240.
15. Callas PW, Pastides H, Hosmer DW: Empirical comparisons of proportional hazards, Poisson, and logistic regression modeling of occupational cohort data. Am J Ind Med 1998, 33:33–47.
16. Breslow N: Covariance analysis of censored survival data. Biometrics 1974, 30:89–99.
17. Therneau TM, Grambsch PM: Modeling survival data: extending the Cox model. statistics for biology and health. New York: Springer; 2000.
18. Kelly PJ, Lim LL: Survival analysis for recurrent event data: an application to childhood infectious diseases. Stat Med 2000, 19:13–33.
19. Rabe-Hesketh S, Skrondal A: Multilevel and longitudinal modelling using Stata. 2nd edition. Stata Press, 4905 Lakeway Drive, College Station, Texas; 2008.
20. Friedman M: Piecewise exponential models for survival data with covariates. Ann Stat 1982, 10:101–103.
21. R Development Corp Team: R: a language and environment for statistical computing. Vienna: R Fundation for Statistical Computing, 2003.
22. StataCorp: Stata statistical software: release 12, College station. Texas: StataCorp MP; 2011.