BMC
Medical Research Methodology

Empirical study of correlated survival times for recurrent events with proportional hazards margins and the effect of correlation and censoring

Rodrigo Villegas[1]*, Olga Julià[2] and Jordi Ocaña[3]

Abstract

Background: In longitudinal studies where subjects experience recurrent incidents over a period of time, such as respiratory infections, fever or diarrhea, statistical methods are required to take into account the within-subject correlation.

Methods: For repeated events data with censored failure, the independent increment (AG), marginal (WLW) and conditional (PWP) models are three multiple failure models that generalize Cox's proportional hazard model. In this paper, we revise the efficiency, accuracy and robustness of all three models under simulated scenarios with varying degrees of within-subject correlation, censoring levels, maximum number of possible recurrences and sample size. We also study the methods performance on a real dataset from a cohort study with bronchial obstruction.

Results: We find substantial differences between methods and there is not an optimal method. AG and PWP seem to be preferable to WLW for low correlation levels but the situation reverts for high correlations.

Conclusions: All methods are stable in front of censoring, worsen with increasing recurrence levels and share a bias problem which, among other consequences, makes asymptotic normal confidence intervals not fully reliable, although they are well developed theoretically.

Background

Cox's proportional hazards model [1] is the most commonly used model for clinical trial data and provides reliable estimates of survival times, as well as the relative risk associated with time-to-event occurrence. As a semi-parametric model, it does not have any constraints on distributional assumptions, which makes it an attractive alternative to parametric models. However, it is not suitable for recurrent event data because survival times in the standard model terminate at the time of the event. In addition, trying to use multiple failure times is inappropriate since events within individuals may be correlated. In this case, the assumption of independence in the standard Cox regression model is violated, introducing statistical

complications. To avoid the errors resulting from analyzing correlated repeated events, sometimes repeated events are disregarded and time to first event is only used, which is clearly a suboptimal approach. Multiple-event analytical techniques for survival data have been developed during the last decades [2] but, as a result of their complex structure and computational requirements, they have not been commonly applied. Advancement in statistical software in recent years has made these methods more accessible to researchers. As a consequence, their use has gradually increased in areas such as clinical research. Text books like Therneau and Grambsch, [2], describe these models in detail with examples, while research articles, such as that of Kelly and Lim [3], provide a very comprehensive explanation of their application to multiple-failure data, with extensive discussion about assumptions. There are three commonly used statistical

*Correspondence: rvillegas.bios@gmail.com
[1]Universitat de Barcelona, Barcelona, Spain
Full list of author information is available at the end of the article

methods for the analysis of recurrent events: Anderson-Gill model [4], Prentice-William-Peterson model [5] and Wei-Lin and Weissfeld model [6]. However, the benefits and drawbacks of each method are still unclear to the medical researchers, specially under varying strengths of correlation between survival times. The aim of this work is to study, through simulations and a real data set, the differences between the three models under presence of low to moderate correlation in survival times. With respect to similar studies like Kelly and Lim [3], we very approximately control the within individual correlation between consecutive recurrent events and we allow decreasing correlations as long as distance between recurrent events increases. In addition our simulation design takes care of different censoring levels and different numbers of recurrences, which are not fixed at 4 events as in [3]. We briefly describe the models, discuss in more detail the simulation procedure, and discuss the results of fitting the simulated data for the different models. In the Example section, we provide an illustration of the models performance using real data from a study in infants with respiratory illness recurrences.

Methods
Basic notations
It is common to model survival times through the hazard function. The Cox proportional hazard model for right censored survival data specifies the hazard function for the ith individual by

$$h_i(t) = h_0(t) \exp(\beta' Z_i), \tag{1}$$

where t represents time, Z_i is the vector of covariates of the ith individual, β is a vector of regression coefficients and $h_0(t)$ is the so-called baseline hazard function that corresponds to the hazard function for an individual with covariables set to zero. As model (1) is formulated through the hazard function, the simulation of appropriate survival times for this model is not straightforward, the effect of the covariates must be translated from the hazards to the survival time. We need to simulate the survival times since software packages for Cox models require the individual survival time data, not the hazard function. Suppose that there are n individuals and that each individual can experience K failures. Times of failure are subject to right censoring and we consider that censoring is non informative, which means that knowledge of a censoring time for an individual or subject provides no further information about the individual's probability of survival at a future time. Let T_{ij} be the time when the jth failure occurs for the ith subject, measured from the subject's study enrollment, and C_i be the corresponding censoring time for individual i. Since the jth failure time T_{ij} is affected by right censoring we define X_{ij} as the minimum of (T_{ij}, C_i), i.e., X_{ij} equals T_{ij} if the event was observed, and C_i if it is censored. Let

$\delta_{ij} = I(T_{ij} \leq C_i)$ be the indicator function taking the value 1 when $T_{ij} \leq C_i$ and 0 otherwise. For simplicity, in our simulations all censoring times are taken as constant, i.e. $C_i = C$ for all i.

Overview of the multiple failure time models
For multiple events the common approaches that generalize the Cox's framework are the independent increment (AG), marginal (WLW) and conditional (PWP) models. These three models differ essentially in the risk sets.

Andersen and Gill (AG) independent increment model
In the AG model [4] it is assumed that recurrent events are unaffected by earlier events that occurred to the same subject so baseline hazards for all events are common. The assumption of mutual independence of the events within a subject is equivalent to the assumption of independent increments in the counting process inside each individual. Each recurrent event for the ith subject is assumed to follow a proportional hazard model where the hazard function is

$$h_i(t) = h_0(t) \exp(\beta' Z_i(t)).$$

Under this model, the risk of a recurrent event for a subject follows the usual proportional hazards assumption but the rank of recurrence is not taken into account. By defining the risk set indicator $Y_{ij}(t) = I(X_{i,j-1} < t < X_{ij})$, the risk set at time t, represented by $\sum_{ij} Y_{ij}(t)$, includes all subjects under observation regardless of the number of occurrences experienced by each subject. As j increases, fewer individuals are likely to experience the event, but this fact does not affect the risk set due to the fact that it does not depend on j; so the coefficients estimates are stable. The AG model provides more efficient inference for a covariate effect than the plain Cox model for the time to the first event, but requires much stronger assumptions than the other models, such as independence among recurrent event times or common baseline hazards. Also, a robust "sandwich" method can be used in the estimation of standard errors [7].

Prentice, Williams and Peterson (PWP) conditional model
Another model for analyzing recurrent events is the PWP model [5] with time scale measured from the beginning of the study to a specified failure. It assumes that a subject is not at risk for the jth event until he/she has experienced event $j - 1$. This model requires the same assumptions as the Cox model, but, unlike the AG model, it allows the baseline hazard to vary from recurrence to recurrence, the hazard function for the jth event for the ith subject is:

$$h_{ij}(t) = h_{0j}(t) \exp(\beta_j' Z_i(t)).$$

The risk set indicator is the same as in the AG model, $Y_{ij}(t) = I(X_{i,j-1} < t < X_{ij})$, but the risk set at time t

is different for each j: $\sum_i, Y_{ij}(t)$. According to the definition of the risk set, the number of subjects is dramatically decreased as j increases. Stable coefficient estimates cannot be obtained for higher ranks of j. The hazard function at time t for the jth recurrence is conditional to the entire previous failures. This model allows different baseline hazards, therefore, estimations for the current event may be affected by earlier events.

Wei, Lin, and Weissfeld (WLW) marginal model

The WLW model [6] uses, for each j, the distribution of time from the study enrollment to the jth recurrent event according to a Cox model, leaving the dependence structure between the observations within the same subject completely unspecified. For the WLW model the risk set indicator is $Y_{ij}(t) = I(X_{i,j} \geq t)$, and the risk set at time t for the jth recurrence, represented by $\sum_i, Y_{ij}(t)$, includes anyone who has not experienced the jth recurrence at time t. In other words, unlike the PWP model, a subject in the risk set at time t does not necessarily have to experience the $(j-1)^{th}$ recurrence. The hazard function of the marginal model for the jth event for the ith subject is

$$h_{ij}(t) = h_{0j}(t)exp(\beta_j'Z_i(t)).$$

The WLW approach obtains the estimate of the covariate effects from the partial likelihood function under the working assumption of independence and then adjusts the variance estimate empirically. Although observations on each subjects are correlated, the β estimation has been shown to be consistent in the case of two events [8]. As the inverse of the information matrix provides a naive variance-covariance estimation, not correctly estimating the true variance, a robust "sandwich" variance-covariance estimation method is preferred.

In all these models we can add a frailty, that is, a random individual effect to explain the dependence of individual recurrent events. In our simulation study we try to precisely control the degree of correlation between adjacent recurrent event times. Frailty makes difficult such a control, so we have not considered it.

Simulation

We carried out a series of simulations to examine the accuracy of the above multiple failure time models in terms of four factors: different sample sizes, censoring levels, numbers of recurrence events and correlation levels. The sample sizes under consideration were $n = (50, 100, 200, 400)$. Survival times were censored at fixed time determined to yield censoring percentages p of 0%, 15%, 30% and 50%. In other words, previously fixed censoring times $C_p = (C_0 = +\infty, C_{15}, C_{30}, C_{50})$ were established in order to ensure the preceding proportions of subjects experiencing a censure. The maximum number of recurrent events under consideration were $K = $

$(3, 6, 9, 12)$. For subject i, the recurrence times t_{i1}, $t_{i2}, \ldots,$ t_{iK} were generated as correlated Weibull deviates as is explained below. Censoring occurred in subject i if $t_{i1} + t_{i2} + \ldots + t_{iK} > C_p$. In this case he/she experienced j events, $j < K$, if $T_{ij} = t_{i1} + \ldots + t_{ij} \leq C_p$ and $t_{i1} + \ldots + t_{i(j+1)} > C_p$. Note that not always censoring occurred only in the Kth event. The correlation levels between adjacent recurrence times were set at $\rho = (0, 0.10, 0.45, 0.80)$. The 320 simulated scenarios were defined by crossing all levels of these four factors. Each simulation series consisted of the generation of $m = 10000$ random samples, each one representing a survival dataset where half of individuals were considered as "control" and half of individuals were considered as "treatment". This covariate was represented by a binary variable with values 0 and 1. The treatment effect was set at $\beta = -1$. For each simulated dataset we adjusted the WLW, AG and PWP models. We used the following measures of the accuracy of the regression parameter estimates and the gain of robustness from the robust approaches: the true coverage of the 95% confidence interval for the parameter β, the relative bias of the β estimators under the three models and the bias of the "robust" and "naive" variance estimators for all three models. The relative sampling bias is defined as the average bias from the m random samples. That is, if $\hat{\beta}_i$ is the estimate on i^{th} simulated dataset, then

$$relative\ sampling\ bias = \frac{1}{m} \sum_{i=1}^{m} \frac{(\hat{\beta}_i - \beta)}{\beta}$$

To simulate the survival times that meet the proportional hazards assumption we used the algorithm proposed by Feiveson [9]. Let U be a random variable with uniform distribution. We can generate a new random variable Y with distribution function F by using the inverse (or pseudo inverse) function F^{-1} [10] as follows

$$Y = F^{-1}(U).$$

In our case, we are interested in simulating times with Weibull distribution with parameters c and κ, that is, with distribution function $F(y) = 1 - exp(-cy^\kappa)$, for $c > 0$, $\kappa > 0$. Note that, usually, the Weibull parameters are λ and κ with $c = \lambda^\kappa$. Therefore, given $U \sim \text{Uniform}[0,1]$,

$$Y = \left(\frac{-\log U}{c} \right)^{\frac{1}{\kappa}} \tag{2}$$

has the aforementioned Weibull distribution. The hazard function corresponding to Y is given by $h(y) = c\kappa y^{\kappa-1}$. Under the Cox model and for a binary covariate X, the parameter c would be equal to some base value c_0 when $X = 0$ and would become equal to $c_1 = c_0 \exp(\beta)$ when $X = 1$.

Simulation of within-subject dependence

Our goal is to simulate recurrent Weibull failure times with correlation ρ between successive recurrence events in the same individual and with decreasing correlation as the distance between events increases. To incorporate this dependence into our simulated data while still maintaining the marginal Weibull distribution for each observation we will use a transformation of a convenient multivariate normal distribution. We describe this procedure in three steps:

1. For any correlation ρ, choose an appropriate ρ_0 following formula (9) derived in the next section. For each subject, i, $i = 1, \ldots, n$, generate a random variate Z_{i1} from a $N(0, 1)$ distribution. Given z_{i1}, generate a random variate z_{i2} from a normal distribution with mean $\rho_0 z_{i1}$ and variance $1 - \rho_0^2$. Given z_{i2}, generate a random variate z_{i3} from a $N(\rho_0 z_{i2}, 1 - \rho_0^2)$; and so on, until z_{iK}. This ensures that each consecutive pair $(z_{ij}, z_{i,j+1})$ is a realization of $(Z_{ij}, Z_{i,j+1})$ following a bivariate normal distribution with standard marginals, $N(0, 1)$ and correlation ρ_0. Also the correlation between Z_{ij} and $Z_{i,j+\tau}$ decreases as τ increases.

2. Transform Z_{ij} into uniform random variables using the standard normal cumulative distribution function Φ: $U_{ij} = \Phi(Z_{ij})$. Variables U_{ij} have a similar correlation structure.

3. Finally, transform the variables U_{ij} into dependent Weibulls using Equation (2). These Weibulls have again a similar correlation structure, with correlation ρ between adjacent recurrence times.

Computation of the correlation ρ_0

In order to simplify our argumentation we use the logarithm of Weibull times. Using delta method it is easy to see that the correlation between times is approximately equal to the correlation between time logarithms. In terms of the original normally distributed Z_{ij}, the simulated log failure times may be expressed as $g_{ij} = G(Z_{ij})$, where

$$G(z) = \log\left[\frac{-\log \Phi(z)}{c}\right]^{\frac{1}{\kappa}} \tag{3}$$

Using second-order Taylor expansions of g_{ij} and $g_{i,j+1}$ about $z = 0$, it can be shown (see Appendix) that

$$Cov\left(g_{ij}, g_{i,j+1}\right) \approx \left\{G'(0)\right\}^2 \rho_0 + \frac{\left\{G''(0)\right\}^2}{4} \rho_0^2 \tag{4}$$

and

$$Var\left(g_{ij}\right) = Var\left(g_{i,j+1}\right) \approx \left\{G'(0)\right\}^2 + \frac{\left\{G''(0)\right\}^2}{2}. \tag{5}$$

Combining (4) and (5), we find that $Corr\left(g_{ij}, g_{i,j+1}\right)$ is a weighted average of ρ_0 and ρ_0^2; namely,

$$Corr\left(g_{ij}, g_{i,j+1}\right) \approx w\rho_0 + \frac{1}{2}(1 - w)\rho_0^2 \tag{6}$$

where

$$w = \frac{\left\{G'(0)\right\}^2}{\left\{G'(0)\right\}^2 + \left\{G''(0)\right\}^2 / 2}. \tag{7}$$

It can be shown (see Appendix) that w does not depend on the parameters κ or c; specifically,

$$w = \frac{\pi \left(\log 2\right)^2}{\pi \left(\log 2\right)^2 + \left(1 - \log 2\right)^2} \approx 0.94128 \tag{8}$$

Thus, solving (6) for the value of ρ_0 such that $Corr\left(g_{ij}, g_{i,j+1}\right) = \rho$, the solution is

$$\rho_0 = \frac{-w + \sqrt{w^2 + 2\rho\left(1 - w\right)}}{(1 - w)}. \tag{9}$$

Software simulation

All the simulations were performed using the library *survival* and the functions *qnorm* and *runif* from the statistical software R2.15.2, 64 bit version, [11]. The default generator is based on a Mersenne-Twister algorithm.

Results

Table 1 shows results for a simple size of 200 subjects and 15% censoring. For simplicity, we provide only graphical results for two degrees of correlation between adjacent recurrent event times (0.1 and 0.8) and two maximum numbers of possible recurrent events (3 and 9). The complete simulation results corresponding to all 320 simulation scenarios are accessible at the url http://www.ub.edu/stat/recerca/materials/SupplEmpirical.htm.

If only the first event is considered, the relative bias, efficiency and coverage of the standard Cox proportional hazard estimator are adequate, as expected, in all the investigated scenarios. We do not report these results because our interest is focused in recurrent data.

Censoring has a very small influence on the performance of the treatment effect estimators, $\hat{\beta}$, for all models under consideration. The graphics are nearly identical for all levels of censoring so only 15% level is reported.

Figure 1 illustrates the relative bias, Figure 2 the bias of the variance estimators and Figure 3 the true coverage of the 95% confidence interval based on the robust variance estimator. All these figures are organized in rows and columns. Each column corresponds to a correlation level ($\rho = 0.1$ and $\rho = 0.8$) and each row to a maximum number of possible recurrences ($K = 3$ and $K = 9$).

With respect to relative bias, the AG estimator of β is stable for all different sample sizes. The bias grows with

Table 1 Relative bias, naive variance, robust variance and coverage robust for 15% censoring and sample size of 200

ρ	Events	Model	Relative bias	Naive var	Robust var	Coverage
0.1	3	AG	-0.193	0.008	0.006	0.308
		PWP	-0.030	0.010	0.011	0.930
		WLW	0.320	0.009	0.021	0.415
	6	AG	-0.254	0.004	0.003	0.003
		PWP	-0.042	0.006	0.007	0.897
		WLW	0.666	0.005	0.021	0.004
	9	AG	-0.278	0.003	0.002	0.000
		PWP	-0.048	0.005	0.006	0.871
		WLW	0.929	0.004	0.022	0.000
	12	AG	-0.291	0.002	0.001	0.000
		PWP	-0.050	0.004	0.005	0.861
		WLW	1.151	0.003	0.022	0.000
0.45	3	AG	-0.296	0.008	0.007	0.069
		PWP	-0.201	0.009	0.011	0.491
		WLW	0.166	0.009	0.022	0.806
	6	AG	-0.343	0.004	0.004	0.001
		PWP	-0.291	0.005	0.006	0.055
		WLW	0.358	0.005	0.022	0.318
	9	AG	-0.350	0.003	0.003	0.000
		PWP	-0.324	0.004	0.004	0.008
		WLW	0.514	0.003	0.021	0.058
	12	AG	-0.349	0.002	0.002	0.000
		PWP	-0.338	0.003	0.003	0.002
		WLW	0.657	0.003	0.022	0.005
0.8	3	AG	-0.381	0.008	0.008	0.016
		PWP	-0.339	0.009	0.011	0.124
		WLW	0.051	0.009	0.024	0.936
	6	AG	-0.448	0.004	0.005	0.000
		PWP	-0.509	0.005	0.005	0.000
		WLW	0.110	0.004	0.023	0.892
	9	AG	-0.462	0.003	0.004	0.000
		PWP	-0.574	0.003	0.003	0.000
		WLW	0.157	0.003	0.023	0.829
	12	AG	-0.462	0.002	0.004	0.000
		PWP	-0.605	0.002	0.003	0.000
		WLW	0.205	0.002	0.022	0.727

the correlation and the number of possible recurrences. The treatment effect is always underestimated. In all circumstances the naive variance estimator is worse than the robust estimator, but both perform very similarly and adequately. The true coverage of the confidence intervals is in general low and decreases with correlation, number of events and, surprisingly, with sample size.

The estimator of β based on model PWP is nearly unbiased for low correlations and stable with respect to the number of recurrences. When the correlation is high, it underestimates the treatment effect β. This bias does not improve with growing sample sizes. Again, the robust variance is better than the naive variance, but both are very similar and perform adequately, improving with
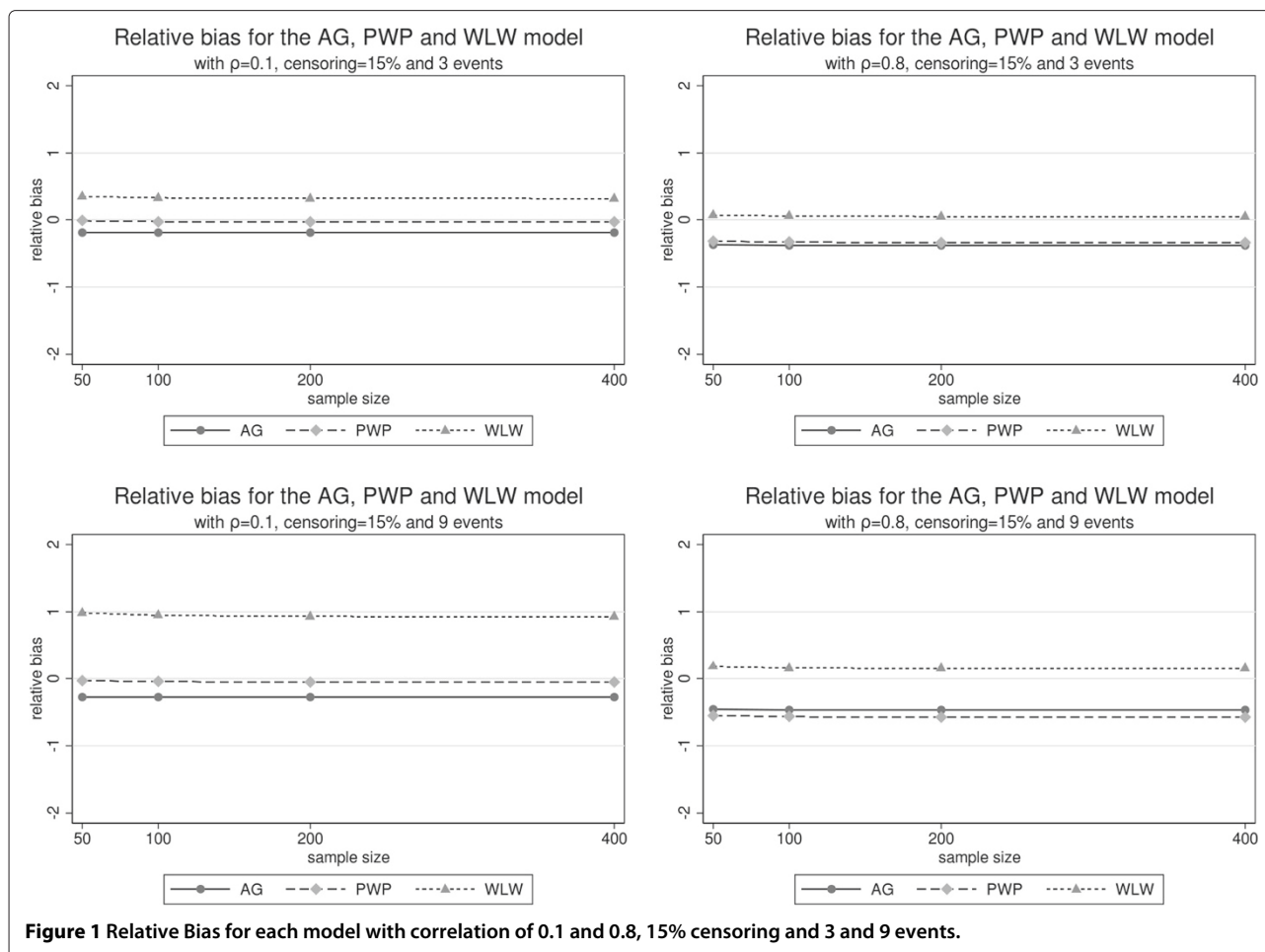
Figure 1 Relative Bias for each model with correlation of 0.1 and 0.8, 15% censoring and 3 and 9 events.

growing sample sizes and correlations. For low correlation levels its coverage is the best, but it quickly falls with increasing correlations and becomes sensitive to growing sample sizes, in the same surprising direction as the AG method.

Under the WLW model β is overestimated. This bias increases with growing recurrence but improves with growing correlation. The naive variance estimator under WLW is very biased, its use should be avoided. The robust variance estimator is clearly better but slightly worse than the other robust estimators for sample sizes below 200. Under low correlation, the confidence interval performs very inadequately, with very low coverages which decrease with growing sample sizes and with growing number of recurrences. On the other hand, its coverage greatly improves with growing correlations, though a tendency to make worse with growing sample sizes and recurrence levels still persists.

Example

The models under consideration are illustrated through the analysis of a cohort of infants with bronchial

obstruction [12]. For this study, 4 months-old children were followed until they reached 12 months of age and recruitment took place during the course of 18 months, from April 1995 to October 1996. The chief goal of the investigation was to estimate the effects of fine particulate matter and wheezing illness in the first year of life. The recurrent events of interest were physician office visits attributable to respiratory illnesses. From the pediatric visit database, a total of 504 infants were eligible for the present study in the pediatric visit database. The events of interest were the times to physician revisit following the initial visit due to examination. Each subject experienced a particular number of visits to physician at various times, which represent the whole observable history of his/her recurrences. For each individual the last time is censored due to the end of the study period. Of these 504 subjects, 475 had at least one revisit during the follow-up time. Table 2 summarizes the number of events experienced by the 475 subjects during the follow-up time. A total of 475 occurrences on revisit were observed, where some infants experienced a rather large number (up to 10 revisits).
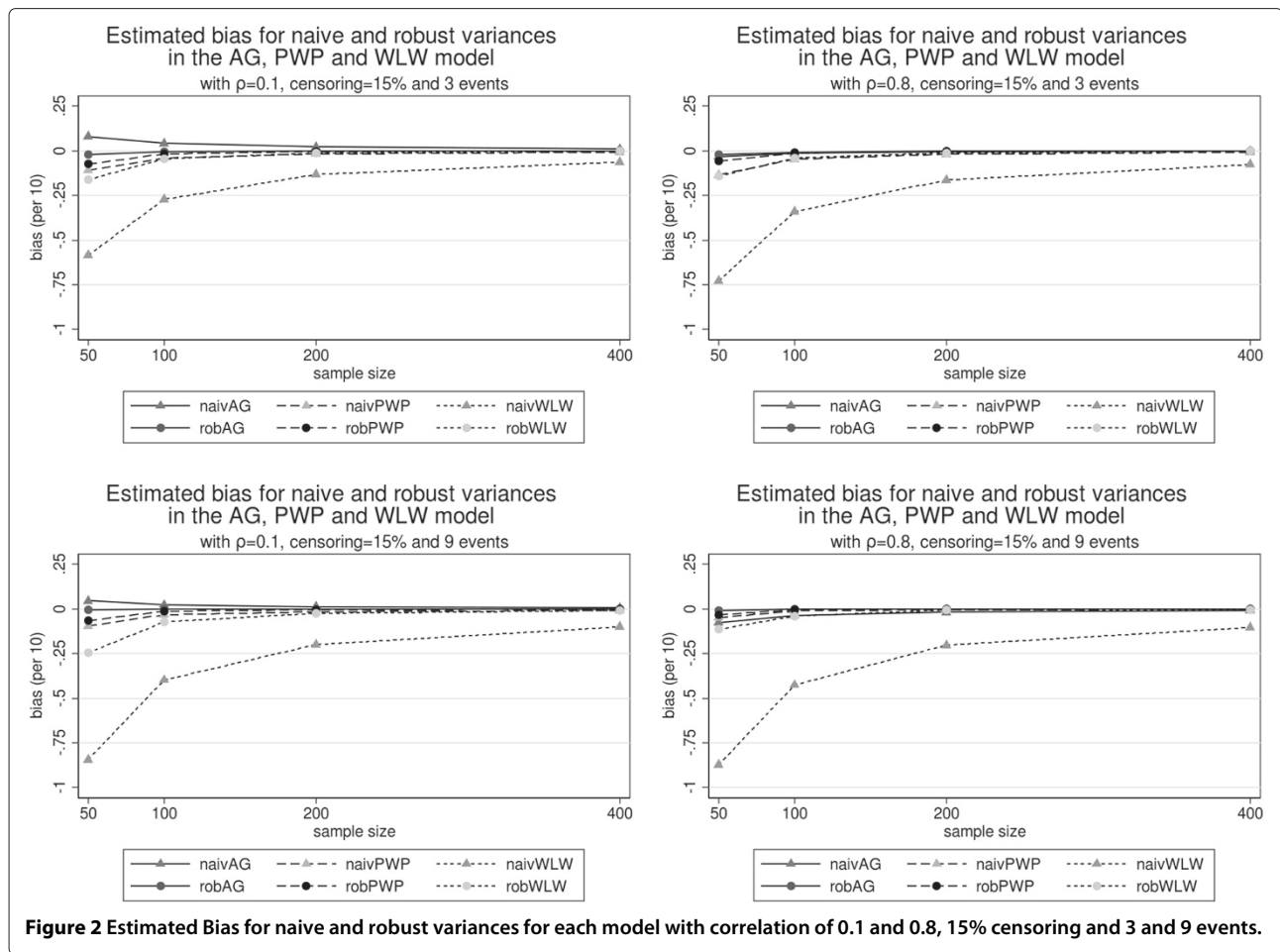
Figure 2 Estimated Bias for naive and robust variances for each model with correlation of 0.1 and 0.8, 15% censoring and 3 and 9 events.
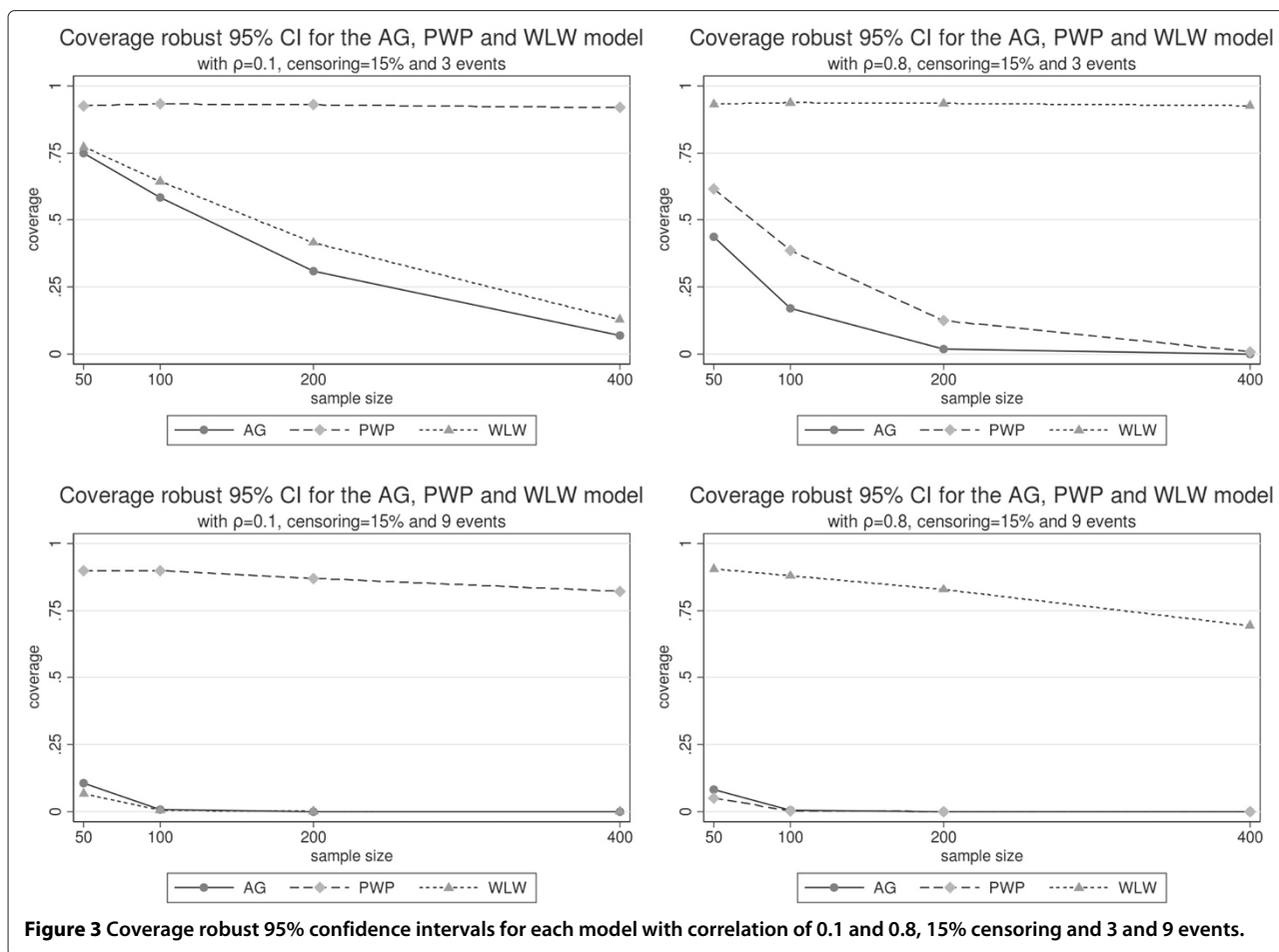
The main purpose of this example is to illustrate the application of recurrent event techniques and provide comparisons between techniques, rather than giving estimates for recurrence of respiratory illnesses. Two characteristics, parents reported smoking (1=yes, 0=no) and family history of asthma (1=yes, 0=no) were included as two univariate analysis in each of the models studied. We chose these two covariates because they are potentially related with the outcome as indicated in the medical literature and also because they are binary variables as we used in our simulation study. Table 3 contains the results of three different models: AG, PWP and WLW, fitted to the respiratory data. The β values estimated by the AG and PWP models are lower than those estimated by the WLW model for both covariates, smoking and asthma. But all of them show a negative effect in the outcome. When we analyze the standard errors, the robust standard errors are larger in all models than those estimated by the naive method. And again, the WLW model has the biggest standard error in comparison to the other models. These results are coherent with the previous simulation results.

Note that the WLW is the only model where asthma is a statistically significant.

Discussion

The simulation results show that each model has different degrees of robustness against variations in the number of recurrent events and correlation. In terms of relative bias, the WLW model shows the worst performance for low correlations but greatly improves for high correlation levels, independently of censoring and sample size. The situation is reversed for the AG and PWP models, which perform better than WLW for low correlation levels. However, these models clearly underestimate the parameter. As some authors have suggested [3], the WLW model overestimates regression coefficients due to a carry-over effect. This is corroborated by our simulation results. Additionally, as we can see from Table 3 in the example, the WLW model presents the highest coefficients and robust standard errors.

The most surprising results refer to the unexpected behavior of all methods with growing sample sizes. First,

Figure 3 Coverage robust 95% confidence intervals for each model with correlation of 0.1 and 0.8, 15% censoring and 3 and 9 events.

their bias remains nearly constant. This was so surprising that we repeated the computations with Stata, with identical results to those obtained with the "coxph" R function. Provided that (as is expected) the true variance of all estimators decreases and its robust estimation greatly improves with growing sample sizes (Figure 2), as a consequence the confidence intervals become more and more narrow around a biased value, which produces a disconcerting, at first sight, fall of coverage (Figure 3).

The AG, PWP and WLW models have been previously compared by various authors (e.g., [3,7,13,14]) using real and simulated data, showing that these models often yield different results for the same data set. This is not unexpected, since distinct models address different research questions. For example, in [3] they found that the WLW model overestimates the treatment effect. Also

they show that as the correlation increases the coverage of the 95% confidence interval based on the robust variance decreases, except for WLW.

Conclusion

In this paper we reviewed and compared the principal methods in the analysis of recurrent event data in terms of censoring and within-subject correlation. The example illustrates the differences between the methods and resulting parameter estimators.

No method can be recommended as the best in all circumstances. The AG and PWP approaches are quite robust in front of low levels of within-subject correlation, but the WLW approach should be recommended under the suspicion of high correlation. All methods share a bias problem which makes the confidence intervals based

Table 2 Number of episodes of illness

	Number of events							
	0	1	2	3	4	5	6-10	Total
N. of subjects (%)	260 (54.7)	101 (21.3)	53 (11.2)	21 (4.4)	21 (4.4)	10 (2.1)	9 (1.9)	475 (100)

Table 3 Coefficients and standard errors obtained in two univariate analysis for 2 covariates: parents reported smoking and family history of asthma

Model	Covariate	Beta	Naive S.E.	Robust S.E.	Robust 95% CI
AG	Smoking	0.158	0.059	0.089	-0.016 , 0.332
	Asthma	0.276	0.111	0.155	-0.028 , 0.579
PWP	Smoking	0.147	0.060	0.088	-0.025 , 0.319
	Asthma	0.251	0.113	0.148	-0.039 , 0.540
WLW	Smoking	0.162	0.058	0.102	-0.037 , 0.362
	Asthma	0.407	0.111	0.189	0.036 , 0.778

on asymptotic normal theory not fully accurate. On the other hand, the robust estimation of the variance of the treatment effect estimator is very reliable, and should be recommended instead of the naive variance estimation, especially for the WLW method.

Appendix

Remember that $\rho(Z_{ij}, Z_{ik}) = \rho_0 > 0$ and $\rho(g_{ij}, g_{i,j+1}) = \rho > 0$ with $g_{ij} = G(Z_{ij})$ where G is defined in (3). To find the correlation between g_{ij} and $g_{i,j+1}$ we will use a second order Taylor expansion (delta method, [15]) for the function G about $z = 0$ as follows:

$$g_{ij} = G(Z_{ij}) = G(0) + G'(0)(Z_{ij} - 0) + \frac{G''(0)}{2}(Z_{ij} - 0)^2 + \cdots$$

$$\approx G(0) + G'(0)Z_{ij} + \frac{G''(0)}{2}Z_{ij}^2$$

since $Z_{ij} \sim N(0,1)$, we have $E(Z_{ij}) = 0$ and $Var(Z_{ij}) = E(Z_{ij}^2) = 1$, then,

$$E(g_{ij}) \approx G(0) + \frac{G''(0)}{2}$$

$$g_{ij} - E(g_{ij}) \approx G'(0)Z_{ij} + \frac{G''(0)}{2}(Z_{ij}^2 - 1)$$

$$g_{i,j+1} - E(g_{i,j+1}) \approx G'(0)Z_{ik} + \frac{G''(0)}{2}(Z_{ik}^2 - 1)$$

and

$$(g_{ij} - E(g_{ij}))(g_{i,j+1} - E(g_{i,j+1})) \approx (G'(0))^2 Z_{ij}Z_{ik}$$
$$+ \frac{G'(0)G''(0)}{2}Z_{ij}(Z_{ik}^2 - 1) + \frac{G'(0)G''(0)}{2}Z_{ik}(Z_{ij}^2 - 1)$$
$$+ \frac{(G''(0))^2}{4}(Z_{ij}^2 - 1)(Z_{ik}^2 - 1).$$

By applying now the $E(\cdot)$ operator taking into account that

$$(Z_{ij}, Z_{ik}) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix} \right),$$

so $E[Z_{ij}Z_{ik}] = \rho_0$ and $E[Z_{ij}(Z_{ik}^2 - 1)] = E[Z_{ik}Z_{ij}^2] = 0$, we have:

$$Cov(g_{ij}, g_{i,j+1}) \approx (G'(0))^2 E(Z_{ij}Z_{ik}) + \frac{(G''(0))^2}{4}$$
$$\times E\left[(Z_{ij}^2 - 1)(Z_{ik}^2 - 1) \right]$$
$$= (G'(0))^2 \rho_0 + \frac{(G''(0))^2}{4}\rho_0^2.$$

For the $Var(g_{ij})$ we proceed in the same way as before,

$$Var(g_{ij}) \approx Var\left(G(0) + G'(0)Z_{ij} + \frac{G''(0)}{2}Z_{ij}^2 \right)$$
$$= (G'(0))^2 Var(Z_{ij}) + \frac{1}{4}(G''(0))^2 Var(Z_{ij}^2)$$
$$= (G'(0))^2 + \frac{1}{2}(G''(0))^2.$$

because $Cov(Z_{ij}, Z_{ik}^2) = 0$, and $Var(Z_{ij}^2) = 2$. Finally the formula for the approximated correlation between g_{ij} and $g_{i,j+1}$ is

$$Corr(g_{ij}, g_{i,j+1}) = \frac{Cov(g_{ij}, g_{i,j+1})}{\sqrt{Var(g_{ij})Var(g_{i,j+1})}}$$
$$= \frac{(G'(0))^2 \rho_0 + \frac{1}{4}(G''(0))^2 \rho_0^2}{(G'(0))^2 + \frac{1}{2}(G''(0))^2}$$
$$= \frac{(G'(0))^2}{(G'(0))^2 + \frac{1}{2}(G''(0))^2}\rho_0$$
$$+ \frac{1}{4}\frac{(G''(0))^2}{(G'(0))^2 + \frac{1}{2}(G''(0))^2}\rho_0^2$$
$$= w \cdot \rho_0 + \frac{(1-w)}{2}\rho_0^2$$

where, $w = \frac{(G'(0))^2}{(G'(0))^2 + \frac{1}{2}(G''(0))^2}$. In order to compute w, we use that:

$$\Phi(z) = \int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \quad \Phi(0) = \frac{1}{2}$$
$$\Phi'(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \quad \Phi'(0) = \frac{1}{\sqrt{2\pi}}$$
$$\Phi''(x) = -\frac{x}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \quad \Phi''(0) = 0.$$

Then,

$$G'(z) = \frac{1}{k} \frac{\Phi'(z)}{\Phi(z) \log \Phi(z)}$$

$$G'(0) = -\frac{\sqrt{2}}{k(\log 2)\sqrt{\pi}}$$

$$G''(z) = -\frac{1}{k} \frac{(\Phi'(z))^2}{(\log \Phi(z))^2 \Phi(z)^2} + \frac{1}{k} \frac{\Phi''(z)\Phi(z) - (\Phi'(z))^2}{\Phi(z)^2 \log \Phi(z)}$$

$$G''(0) = \frac{2(\log 2 - 1)}{k\pi(\log 2)^2}.$$

And finally,

$$w = \frac{(G'(0))^2}{(G'(0))^2 + \frac{1}{2}(G''(0))^2} = \frac{\pi(\log 2)^2}{\pi(\log 2)^2 + (\log 2 - 1)^2}.$$

Note that w does not depend neither on k nor on c.

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

Authors' contributions
RV conducted literature review, wrote computer programs, produced graphs and contributed to the mathematical development. OJ developed the mathematical framework and derived mathematical results. JO contributed to the improvement of simulation techniques and software development. All authors prepared the manuscript and have read and approved the final version.

Author details
[1]Universitat de Barcelona, Barcelona, Spain. [2]Departament de Probabilitat, Lògica i Estadística, Universitat de Barcelona, Barcelona, Spain. [3]Departament d'Estadística, Universitat de Barcelona, Barcelona, Spain.

References
1. Cox DR: **Regression models and life-tables (with discussion).** *J R Stat Soc Ser B* 1972, **34**(2):187–220.
2. Therneau TM, Grambsch PM: *Modeling Survival Data: Extending the Cox Model.* New York: Springer; 2000.
3. Kelly PJ, Lim LL: **Survival analysis for recurrent event data: an application to childhood infectious diseases.** *Stat Med* 2000, **19**:13–33.
4. Anderson PK, Gill RD: **Cox's regression model for counting processes: a large sample study.** *Ann Stat* 1982, **10**:1100–1120.
5. Prentice RL, Williams BJ, Peterson AV: **On the regression analysis of multivariate failure time data.** *Biometrika* 1981, **68**:373–379.
6. Wei LJ, Lin DY, Weissfeld L: **Regression analysis of multivariate incomplete failure time data by modeling marginal distributions.** *J Am Stat Assoc* 1989, **84**(408):1065–1073.
7. Therneau TM, Hamilton SA: **rhDNase as an example of recurrent event analysis.** *Stat Med* 1997, **16**:2029–2047.
8. Lin DY: **Cox regression analysis of multivariate failure time data.** *Stat Med* 1994, **13**:2233–2247.
9. Feiveson AH: **Power by simulation.** *Stata J* 2002, **2**:107–124.
10. Hoel P, Port S, Stone C: *Introduction to Probability Theory.* Boston: Houghton Mifflin Company; 1971.
11. R Core Team: *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing; 2013. [http://www.R-project.org/]. [ISBN 3-900051-07-0].
12. Pino P, Walter T, Oyarzun M, Villegas R, Romieu I: **Fine particulate matter and wheezing illnesses in the first year of life.** *Epidemiology* 2004, **15**(6):702–708.
13. Gao S, Zhou XH: **An empirical comparison of two semi-parametric approaches for the estimation of covariate effects from multivariate failure time data.** *Stat Med* 1997, **16**:2049–2062.
14. Wei LJ, Lin DY, V G D: **An overview of statistical methods for multiple failure time data in clinical trials.** *Stat Med* 1997, **16**:833–839.
15. Rice J: *Mathematical Statistics and, Data Analysis.* 2nd edition. Belmont: Duxbury Press; 1995.