International Journal of Epidemiology, 2015, 334–344
doi: 10.1093/ije/dyu262
Advance Access Publication Date: 19 January 2015
Original article

Education Corner

Joint modelling of repeated measurement and time-to-event data: an introductory tutorial

Özgür Asar,[1]* James Ritchie,[2] Philip A Kalra[2] and Peter J Diggle[1,3]

[1]CHICAS, Lancaster Medical School, Lancaster University, Lancaster, UK, [2]Vascular Research Group, Manchester Academic Health Sciences Centre, University of Manchester, Salford Royal NHS Foundation Trust, UK and [3]Institute of Infection and Global Health, University of Liverpool, Liverpool, UK

*Corresponding author. CHICAS, Lancaster Medical School, Faculty of Health and Medicine, Lancaster LA1 4YG, UK. E-mail: o.asar@lancaster.ac.uk

Accepted 2 December 2014

Abstract

Backgound: The term 'joint modelling' is used in the statistical literature to refer to methods for simultaneously analysing longitudinal measurement outcomes, also called *repeated measurement* data, and time-to-event outcomes, also called *survival* data. A typical example from nephrology is a study in which the data from each participant consist of repeated estimated glomerular filtration rate (eGFR) measurements and time to initiation of renal replacement therapy (RRT). Joint models typically combine linear mixed effects models for repeated measurements and Cox models for censored survival outcomes. Our aim in this paper is to present an introductory tutorial on joint modelling methods, with a case study in nephrology.

Methods: We describe the development of the joint modelling framework and compare the results with those obtained by the more widely used approaches of conducting separate analyses of the repeated measurements and survival times based on a linear mixed effects model and a Cox model, respectively. Our case study concerns a data set from the Chronic Renal Insufficiency Standards Implementation Study (CRISIS). We also provide details of our open-source software implementation to allow others to replicate and/or modify our analysis.

Results: The results for the conventional linear mixed effects model and the longitudinal component of the joint models were found to be similar. However, there were considerable differences between the results for the Cox model with time-varying covariate and the time-to-event component of the joint model. For example, the relationship between kidney function as measured by eGFR and the hazard for initiation of RRT was significantly underestimated by the Cox model that treats eGFR as a time-varying covariate, because the Cox model does not take measurement error in eGFR into account.

Conclusions: Joint models should be preferred for simultaneous analyses of repeated measurement and survival data, especially when the former is measured with error and the association between the underlying error-free measurement process and the hazard for survival is of scientific interest.

Key words: Chronic kidney disease, cohort study, epidemiology, joint modelling of longitudinal and survival data, measurement error, medical statistics, statistical software

Key Messages

- Longitudinal studies often include both repeated measurement and survival outcomes. Common practice is to analyse these data separately. This is mostly due to the lack of awareness of the available tools for simultaneous analysis.
- Measurement error in a time-varying covariate biases the estimate of the underlying association with the hazard for survival towards zero. Joint modelling corrects this.
- Joint modelling of longitudinal and survival data is preferable to separate analyses, both to make optimal use of the available information and to obtain unbiased estimates of the model parameters.
- The availability of publicly available software to fit joint models and examples on their use will encourage wider use of joint models.

Introduction

Prospective medical studies typically record a variety of covariates on each subject, some fixed (e.g. gender), others time-varying (e.g. age), together with two fundamentally different kinds of outcome: longitudinal data at a regular or irregular sequence of time-points, also called *repeated measurements* (e.g. estimated glomerular filtration rate, eGFR); and time-to-event outcomes, also called *survival data* [e.g. time to initiation of renal replacement therapy (RRT)].

Repeated measurement and survival data require different statistical methods, and are traditionally analysed separately. Typical properties of these data are: (i) repeated measurement sequences are intermittently collected and subject to measurement error; (ii) occurrence of the survival event terminates the underlying measurement process, potentially in an informative manner; and (iii) the underlying measurement process affects the hazard for survival. Together, these properties imply that separate analysis of repeated measurement and survival outcomes is potentially inefficient, because it does not fully exploit the dependence between the repeated measurement process and the hazard for survival, and leads to biased estimation of the association between the two, because it ignores measurement error.

For example, in nephrology it is important to understand the relationship between changes in a patient's renal function over time and the corresponding changes in their survival prognosis; but neither the changes in renal function nor the hazard for RRT are directly observable at all times. For this reason, we need to build a statistical model that relates these unobservable quantities to each other and to the observable data. These data consist of the intermittently measured, error-prone and possibly informatively censored eGFR measurements and the observed, possibly censored, times to RRT for each patient in the study.

Statistical methods for repeated measurement and survival data have generated extensive, but largely separate, literatures. For book-length reviews, see for example Diggle *et al.* (2002)[1] or Fitzmaurice *et al.* (2011)[2] for the former, and Kalbfleisch and Prentice (2002)[3] or Kleinbaum and Klein (2012)[4] for the latter.

Recently, simultaneous analysis of these two types of data has become possible through the development of the so-called *joint modelling* methods: see, for example, Wulfsohn and Tsiatis (1997),[5] Henderson *et al.* (2000),[6] Diggle *et al.* (2007)[7] and Rizopoulos (2012).[8] Much of the early methodological work was stimulated by problems arising in AIDS research.[5,9] More recently, joint modelling methods have been adopted in other areas of clinical research, including cancer,[10] cardiovascular disease[11] and kidney transplantation studies.[12,13] However, joint modelling methods remain under-used, and the absence of an accessible introduction in the epidemiological literature inhibits their wider adoption. The aim of this paper is to provide an introductory tutorial on joint modelling embedded in a specific application in nephrology and including an illustration of open-source software for joint modelling that is available within the R[14] computing environment.

The paper is organised as follows. We first provide details of the data set that we use throughout the paper, and define the required statistical terminology. We then formulate repeated measurement, survival and joint models for these data and discuss their basic properties.

We then use the models to investigate the effect of changes in kidney function on the hazard for RRT, and discuss our findings. R scripts to reproduce our analyses are provided in the online supplementary material (available as Supplementary data at *IJE* online).

Materials and Methods

Patient population

Patients were selected from the Chronic Renal Insufficiency Standards Implementation Study (CRISIS[15,16]) run by Salford Royal NHS Foundation Trust (SRFT). CRISIS is an ongoing prospective observational study of outcome in all-cause chronic kidney disease (CKD) that has continued to recruit patients since 1 October 2002. Patient records are updated at annual nephrology follow-ups by trained research nurses. Renal function is estimated using the four-variable MDRD equation:[17]

$$\text{eGFR} = 175 \times \left(\frac{\text{SCr}}{88.4}\right)^{-1.154} \times \text{age}^{-0.203} \times 0.742\ \text{I(female)} \times 1.21\ \text{I(black)}, \qquad (1)$$

where SCr denotes serum creatinine. In our analysis, we ignored the ethnicity term in Equation 1, because most of the patients in the data set we analysed were Caucasian (96.3%). Predefined study end-points are death (confirmed by the Office for National Statistics) and initiation of RRT, defined as chronic haemodialysis, peritoneal dialysis or transplantation. In this paper, we consider data collected until 30 July 2012 and initiation of RRT as the survival outcome. There are 1611 patients with a total number of 3154 follow-up measurements.

Explanation of statistical terms

Repeated measures are eGFR measurements belonging to the same patient but performed at different times, here corresponding to hospital visits. *Measurement times* are the follow-up times at each hospital visit, defined as the years elapsed between study entry and hospital visits. *Measurement error* is the difference between the computed value of eGFR and the true (isotopic) GFR. A *time-constant* or *baseline* covariate is one whose value does not change over time, e.g. gender, whereas a *time-varying* covariate is one whose value does change over time and is available at all times, e.g. age. Covariates are to be regarded as *inputs* to a biomedical system, whereas outcome variables, here repeated measurements of eGFR and time to initiation of RRT, are to be regarded as *outputs*.

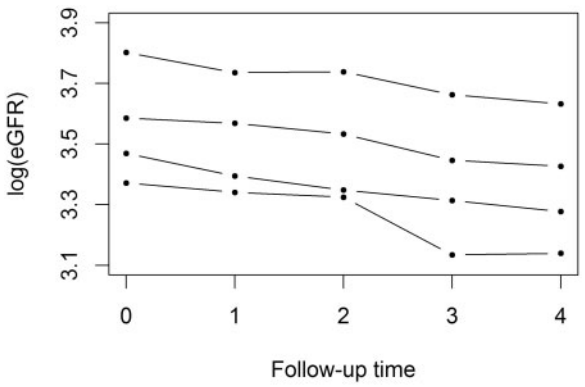

Figure 1. Hypothetical longitudinal data for four patients with five follow-ups.

A *survival* outcome is the time, from a defined origin, at which an event of clinical interest (e.g. initiation of RRT) occurs. Typically, survival times T, can be either *observed* or *censored*, the latter meaning that observation of the subject in question is terminated before the event of clinical interest occurs; hence the data tell us that T is at least T_0, but we do not know the exact value of T. In our example, patients who had either died or were still alive but had not begun RRT by 30 July 2012 are censored for initiation of RRT, and we know only that initiation of RRT happened, if at all, after 30 July 2012. Censoring is *non-informative* if it is statistically independent of the outcome of interest, *informative* otherwise. In our example, censoring due to death could be informative or non-informative, depending on the cause of death, whereas censoring at the study end-date, 30 July 2012, is unambiguously non-informative.

Finally, a *random effect* is a patient-specific coefficient that represents between-patient heterogeneity in an outcome variable that cannot be explained by measured covariates. This is best understood through an example. Figure 1 shows hypothetical data on eGFR measured annually over 5 years for four different patients. All four patients show an approximately linear decrease in eGFR over time, but from clearly different initial values. A simple mathematical representation of this might be:

$$Y_{ij} = A_i + \beta * t_{ij} + Z_{ij},$$

where Y_{ij} is the j^{th} $(j = 1, \ldots, 5)$ measurement for subject $i\,(i = 1, \ldots, 4)$ at time t_{ij} $(t_{ij} = 0, \ldots, 4)$ and Z_{ij} is the corresponding (random) measurement error. The slope, β, is the same for all patients whereas the values of A_i differ among patients. In treating A_i as a random effect, we are assuming that its values are drawn from a statistical distribution. If patients also differed in their rate of decrease in eGFR, we would replace the *fixed effect* β by a random effect, B_i, again assumed to be drawn from a statistical

distribution. A useful way to think about random effects is as proxies for unmeasured covariates.

Rationale for joint modelling

The distinction between covariates and outcome variables is an operational one, in the sense that the same biological construct may be regarded as an input or an output in different studies. For example, in renal research we could envisage a study whose primary objective was to investigate the relationship between blood pressure and the hazard for end-stage renal failure. In that context, blood pressure would be an input, and progression to end-stage renal failure an output. However, we could equally envisage a study of the efficacy of an antihypertensive medication in which dose would be an input and blood pressure an output. A second, more technical distinction is that in order to obtain an unbiased estimate of the effect of a covariate on an outcome variable using standard survival analysis methods, it is necessary that the covariate can be measured at all times and without error.[18] Hence in the present context, if our only objective was to understand the effect of renal function on the hazard for initiation of RRT *and* we were able to monitor error-free GFR continuously over time, we would treat GFR as a time-varying covariate and formulate a statistical model for a patient's hazard given their GFR history. But this is infeasible. We therefore need to model eGFR measurements and time to initiation of RRT jointly in order to understand the relationship between the underlying error-free GFR and the hazard for initiation of RRT. This is shown schematically in Figure 2.

The two components of the resulting joint model, which we explain in detail in the next section of the paper, are:

i. a linear mixed model for the time-course of eGFR;
ii. and a proportional hazards model for the time to initiation of RRT with time-varying random effects.

Both components of the joint model will include terms for measured covariates and unmeasured, error-free GFR, which we treat as a time-varying, patient-specific random effect, GFR(t). This framework, and in particular the linkage of the two components of the joint model through a shared random effect, allow us to answer a range of questions simultaneously according to the goals of each specific application. For example:

i. what is the typical pattern of progression in GFR, and how is this affected by baseline or time-varying covariates;
ii. and how do changes in level of GFR affect survival prognosis?

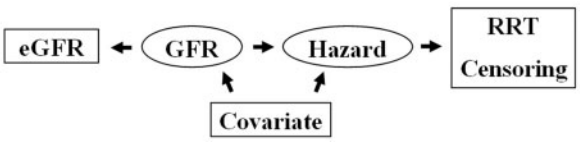

Figure 2. The underlying mechanism of the longitudinal and survival processes. Rectangles denote observed outcomes, ellipses unobserved quantities and arrows directed statistical dependencies. The causal chain of interest runs from GFR to hazard for RRT, whereas eGFR does not 'cause' RRT but is statistically related to RRT through its dependence on the unobserved GFR.

Links between missing data mechanisms and joint modelling

The term 'missing data' refers to data that are intended to be observed, but are unobserved for some reason.[19–21] This is a common occurrence in both longitudinal and cross-sectional studies, either in the explanatory variables or the outcome variables or both. Missing data in longitudinal studies can be either *intermittent*, i.e. a patient might miss some of their hospital visits and return to the study later, or *drop-out*, i.e. a patient might leave the study prematurely. Missing data can be: (i) missing completely at random; (ii) missing at random; or (iii) missing not at random (MNAR). Details of these mechanisms can be found in any material on missing data; see for example, Little and Rubin (2002),[19] Molenberghs and Kenward (2007)[20] and Ibrahim and Molenberghs (2009).[21] There is a close link between drop-out and joint modelling in that drop-out time can be considered as a survival outcome. An operational distinction is that, in the missing data literature, drop-out is typically inferred from a patient's failure to present at a scheduled follow-up time and treated as a discrete-time outcome whereas, in the joint modelling literature, event-time of interest is either recorded exactly or right-censored at the study end-time. The MNAR case is of particular interest in the present context, because it implies that the drop-out time conveys information about the unobserved, error-free longitudinal measurement process over and above that provided by the longitudinal measurements that are observed before drop-out. An example of MNAR in the context of nephrology would be a randomised clinical trial in which patients are likely to drop out when they perceive a lack of benefit from the diet.

Explanation of statistical models

Repeated measurements

The most widely used class of models for repeated measurement data is the linear mixed effects model.[22] This is

defined by:

$$Y_{ij} = Y_i^*(t_{ij}) + Z_{ij} = X_{ij}\beta + W_{ij}B_i + Z_{ij}. \quad (2)$$

Here, Y_{ij} denotes the j^{th} eGFR measurement for the i^{th} patient, a value which is typically measured with error, $Y_i^*(t)$ denotes the true GFR level at time t, and Z_{ij} denotes measurement error. A patient's true GFR level can be decomposed into two components: fixed effects $X_{ij}\beta$; and random effects, $W_{ij}B_i$. The fixed effects represent the expected behaviour of kidney function, averaged over all patients who share the same covariate information; hence, X_{ij} is a vector containing the values of covariates that relate to the i^{th} patient at the time of their j^{th} eGFR measurement. The effects of changing the values of the covariates are represented by the corresponding elements of the regression parameter vector β. The random effects describe how patient-specific true GFR levels deviate from their expected behaviour. Each W_{ij} is a vector containing the values of covariates that relate to the i^{th} patient at the time of their j^{th} eGFR measurement, whereas B_i is analogous to β but, rather than taking a fixed unknown value, varies randomly among patients. Typically the B_i are assumed to follow a zero mean multivariate Normal distribution. There is no requirement for the same covariates to be included in the fixed and random effect components of the model, but typically the latter is a subset of the former. Both X_{ij} and W_{ij} can include time-constant and time-varying covariates.

Parameters are typically estimated by maximum or restricted maximum likelihood (ML and REML, respectively), and the random effects B_i are predicted by their conditional expectations given the data, so as to minimize mean square prediction error. ML is a general method for estimating parameters in complex statistical models that is known to have desirable theoretical properties (see for example, Pawitan 2001[23]). Many familiar elementary statistical methods can be derived as special cases of ML or REML estimations including, for example, t-tests, linear regression and generalized linear modelling. Note that no survival information is considered in Equation 2.

Survival times

The most widely used model for analysing survival data is the Cox proportional hazards model.[24] This is given by:

$$\lambda_i(t|K_i) = \lambda_0(t)\exp(K_i\alpha). \quad (3)$$

Here, $\lambda_i(t|K_i)$ is the hazard for the i^{th} patient to experience the event of interest, e.g. initiation of RRT, at time t. It depends on *time-constant* covariates represented by the elements of a vector of covariates K_i with associated regression parameters α, and a baseline hazard function $\lambda_0(t)$ that represents the hazard for (possibly hypothetical) patients all of whose covariates take the value zero. In most applications, the main interest is in estimating α, which describes how the covariates affect the relative, rather than absolute, hazard. An attractive feature of the Cox proportional hazards methodology is then that it allows the baseline hazard to be left unspecified. If the baseline hazard is of interest, it can be estimated non-parametrically, or modelled parametrically with a specified class of lifetime distributions.[3] A common choice for a parametric specification is the Weibull hazard, which follows a power law, $\lambda_0(t) = \lambda_0 k t^{k-1}$ where $\lambda_0 > 0$ and $k > 0$. Piecewise constant or regression spline function are popular choices for $\lambda_0(t)$ if more flexible parametric modelling is required.[8] Estimates of α are typically obtained by maximizing the partial likelihood[24,25] in the Cox model, or by ML estimation in parametric models.

In principle, time-varying covariates can be added to the model given in Equation 3 by making the elements of K_i functions of time, hence $K_i(t)$; the resulting model is known as the Cox model with time-varying covariate. However, this requires all time-varying covariates to be measured continuously and without measurement error, which is only feasible in special cases, for example where the elements of $K_i(t)$ are functions of time itself. Note however, that in the Cox model any function of time alone is absorbed into the baseline hazard, $\lambda_0(t)$, and therefore cannot be estimated using the partial likelihood. In some applications, continuous measurement of time-varying covariates is induced by interpolating between actual measurements, but this is both an artificial device and also takes no account of measurement error. A key advantage of joint modelling is its ability to handle irregularly and imperfectly measured time-varying covariates correctly.

Inclusion of random effects in the model given in Equation 3 is comparatively straightforward. The simplest example takes the form:

$$\lambda_i(t|K_i, A_i) = \lambda_0(t)\exp(K_i\alpha + A_i), \quad (4)$$

where the A_i are drawn from a statistical distribution. In the survival literature, the quantity $\exp(A_i)$ is called the *frailty* of the i^{th} patient, and the distribution of the A_i is scaled so that the average frailty is 1.

Joint modelling

A joint model for data on eGFR and time to initiation of RRT can now be defined by the following two equations:

$$Y_{ij} = Y_i^*(t_{ij}) + Z_{ij} = X_{ij}\beta + W_{ij}B_i + Z_{ij}, \quad (5)$$

$$\lambda_i\left(t|K_i, Y_i^*(t)\right) = \lambda_0(t)\exp\left(K_i\gamma_1 + Y_i^*(t)\gamma_2\right). \quad (6)$$

Here, γ_2 measures the relationship between the unmeasured, error-free GFR process, $Y_i^*(t)$, and the time to initiation of RRT. The fundamental feature of joint modelling is that repeated measurement and survival data are modelled simultaneously. The algorithm for estimating the parameters of the joint model given in Equations 5 and 6 also exploits the model assumptions to predict the values of $Y_i^*(t)$ at all times t, and thereby to estimate the associated regression parameter γ_2 while making proper allowance for the measurement error in the observed eGFR values. The model can be extended to include particular features of the error-free GFR processes $Y_i^*(t)$. For example, rate of change in GFR can be added to Equation 6 as an additional term with its own regression parameter, i.e. $Y_i^{*'}(t)\gamma_3$. Other possible extensions include interactions of kidney function with a set of baseline covariates; lagged or cumulative effect of kidney function on the hazard for survival. Alternatively, only the random effect component of the longitudinal sub-model might be included in the survival sub-model instead of the current kidney function level, $Y_i^*(t)$. For further details, see Chapter 5 of Rizopoulos (2012)[8] and Henderson *et al.* (2000).[6]

Parameters of joint models are typically estimated by maximizing the likelihood, and random effects are predicted by their conditional expectations given all of the data. The interpretations of the parameters of a joint model are the same as for their linear mixed effects and Cox components.

The benefits of joint modelling are not cost free. The main disadvantages of joint modelling are the increase in computational effort required to fit the models and the relative scarcity of software to enable their routine use. The former is only a significant problem when dealing with very large data sets, in particular data sets with large numbers of repeated measurements on each subject. The latter is being addressed by the development of packages such as JM[26] and joineR[27] that run within the open-source R computing environment.

Framework for statistical analysis

Since some patients missed their annual data updates, there are intermittent missing values in the data set, which we treated missing at random. Each patient contributed both repeated measurement and survival outcomes; the former are the repeated measurements of kidney function as determined by eGFR, the latter are (possibly censored) times to initiation of RRT. In our analysis, we treated death before initiation of RRT as a right-censored event-time. Our analysis comprised three main steps: (i) separate longitudinal analysis of repeated eGFR measurements; (ii) separate survival analysis of time to initiation of RRT using the Cox model with eGFR treated as a time-varying covariate by carrying forward each observed value of eGFR at a constant level until the next observed value on the same patient; and (iii) joint analysis of the repeated eGFR measurements and time to initiation of RRT. In the first step, we built a linear mixed effects model with repeated eGFR data as the response variable, ignoring the potentially informative nature of the censoring of each eGFR sequence by the occurrence of the survival event. In the second step, we analysed RRT with *observed* repeated eGFR measurements treated as a time-varying covariate. In the third step, we considered joint analysis of repeated eGFR measurements and time to initiation of RRT with the current (unobserved) value of GFR included in the survival sub-model in addition to the baseline covariates. We analysed log-transformed eGFR ($Y = \log(eGFR)$) data throughout for the following three reasons. First, this transformation leads to an approximately linear relationship with age and approximately symmetrical scatter about the long-term trend. Second, analyses of log(eGFR) and log-transformed creatinine would be equivalent when the former is re-adjusted with age and gender (see Equation 1). Finally, using a log transformation leads to an interpretation of the fitted model in terms of relative, rather than absolute, change in eGFR, which relates more directly clinical guidelines for monitoring changes in kidney function.

Results

Study population

All of the log(eGFR) measurements are shown in Figure 3. Individual trajectories for 10 patients randomly selected among those with at least three observations are highlighted. Median age at recruitment was 67.2 years (IQR 55.6-74.9); 603 (37.4%) of the patients were female; 1551 (96.3%) of the patients were Caucasian; 509 (31.6%) had SRFT as the base hospital. Mean log(eGFR) at the first

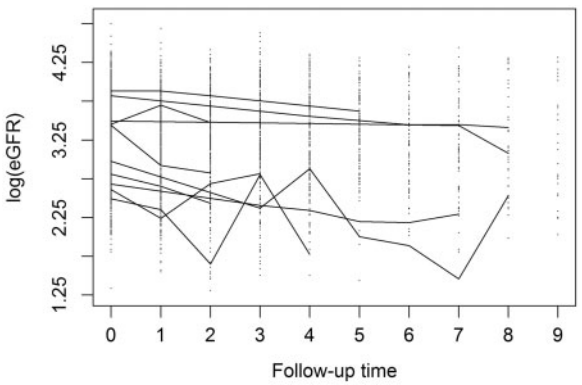

Figure 3. All log(eGFR) measurements. Trajectories for 10 randomly selected patients are shown as connected line-segments.

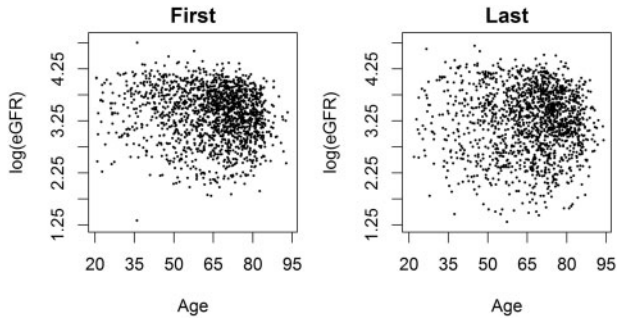

Figure 4. log(eGFR) measurements at the first (left panel) and the last (right panel) follow-ups.

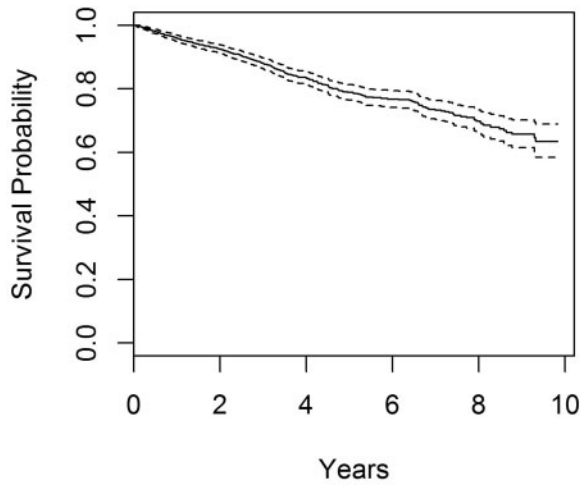

Figure 5. Kaplan-Meier survival plot for RRT as the survival event.

hospital visit was 3.4 ml/min/1.73 m^2 (standard deviation 0.5); and mean log(eGFR) at the last hospital visit was 3.3 ml/min/1.73 m^2 (standard deviation 0.6). The log(eGFR) measurements at first and last visits are displayed in the upper and lower panels of Figure 4, respectively. In total, 516 (32%) of the patients had diabetes mellitus, 1086 (67.4%) were current or former smokers and 342 (21.2%) had a history of coronary artery disease (defined as previous myocardial infarction and/or coronary revascularization procedure). Median follow-up period was 3.9 years (IQR 2.2-5.6); 304 (19%) of the patients experienced initiation of RRT. The Kaplan-Meier survival plot for RRT as the survival outcome is displayed in Figure 5.

Separate analyses

Details of the variables considered in the longitudinal model for log(eGFR) and in the survival model are presented in Table 1. Note that we decompose age at measurement into age at recruitment and time since recruitment (in years) in order to differentiate between cross-sectional and longitudinal effects of age. To explain this distinction,

Table 1. Covariates used in analyses of the CRISIS data set

Variable	Explanation
Baseline age	(Date at study start -date at birth)/365.25
Follow-up	Age at measurement -age at baseline
Hospital base	1 if base hospital is SRFT, 0 otherwise
Gender	1 if male, 0 if female
Smoking	1 if ex or current smoker, 0 if never smoked
Alcohol	1 if alcohol consumer, 0 if abstinent from alcohol
Diabetes	1 if type 1 or type 2 diabetes, 0 if no diabetes
Comorbidity	1 if at least one of myocardial infarction, coronary artery bypass surgery or stenting, 0 otherwise

consider the following simple regression model, in which age_0 denotes age on entry to the study:

$$\log(\text{eGFR}) = \alpha + \beta_1 * \text{age}_0 + \beta_2 * (\text{age} - \text{age}_0) + \text{noise}.$$

In this model, the parameters β_1 and β_2 represent the cross-sectional and longitudinal effects of age, respectively. If $\beta_2 = \beta_1$ the model reduces to:

$$\log(\text{eGFR}) = \alpha + \beta_1 * \text{age} + \text{noise},$$

but there is no reason in general why this should be so. For example, if the stages of the disease for different patients are approximately the same at each follow-up, the cross-sectional effect β_1 would be approximately zero; but as patients typically lose renal function over time, the longitudinal effect β_2 would be negative.

Longitudinal model

We fit the so-called random-intercept-and-random-slope version of the model given in Equation 2, namely:

$$Y_{ij} = Y_i^*(t_{ij}) + Z_{ij} = X_{ij}\beta + A_i + B_i t_{ij} + Z_{ij}, \qquad (7)$$

where t_{ij} denotes time since recruitment. The fixed effects estimates from the separate longitudinal model for change in log(eGFR) are presented in Table 2. Kidney function was found to decrease with increasing age at study start [Estimate = −0.005, 95% confidence interval (CI) −0.007, −0.003] and with increasing time under observation (Estimate = −0.064, 95% CI −0.073, −0.056). Recall that parameter estimates represent relative rather than absolute changes. Hence, a 1-year increase in age at study start was associated with a relative decrease of 0.5% (=(exp(−0.005)−1)*100) in expected eGFR. Similarly, a 1-year increase in time under observation was associated with a relative decrease of 6.2% in expected eGFR. Patients living in the catchment area of SRFT were found to have 15.1% higher expected eGFR at recruitment compared with the patients initially managed at satellite units.

Table 2. Estimated regression parameters, 95% confidence intervals (95% CI), standard errors (SE), *p*-values (p) and percentage relative effects (RE %) in separate longitudinal analysis of the CRISIS data set. The response variable is log-transformed eGFR, RE % corresponding to an estimate $\hat{\beta}$, expressed as expected percentage change in eGFR, calculated as (exp($\hat{\beta}$)−1)*100

Variable	Estimate (95% CI)	SE	*p*	RE %
Intercept	3.585 (3.463, 3.707)	0.062	<0.001	NA
Baseline age (years)	−0.005 (−0.007, −0.003)	0.001	<0.001	−0.5
Follow-up (years)	−0.064 (−0.073, −0.056)	0.004	<0.001	−6.2
Hospital base	0.141 (0.088, 0.194)	0.027	<0.001	+15.1
Gender	0.080 (0.028, 0.133)	0.027	0.003	+8.3
Smoking	0.014 (−0.040, 0.068)	0.028	0.601	+1.4
Alcohol	0.043 (−0.008, 0.093)	0.026	0.098	+4.4
Diabetes	−0.097 (−0.151, −0.044)	0.027	<0.001	−9.2
Comorbidity	−0.024 (−0.087, 0.040)	0.032	0.468	−2.4

Males were found to have 8.3% higher expected eGFR than females. Patients who had type 1 or type 2 diabetes were found to have 9.2% lower expected eGFR than non-diabetic patients. No differences in expected eGFR were found between ex or current smokers and non-smokers, between alcohol consumers and abstainers from alcohol or between patients who did or did not have at least one comorbidity event; respective *p*-values were 0.601, 0.098 and 0.468.

Survival model

Results obtained from a Cox model for the survival outcome are displayed in Table 3. Time-varying eGFR was associated with the hazard for RRT. The corresponding parameter estimate was −2.510 (95% CI −2.691, −2.330); hence, a 1% reduction in eGFR was associated with a 2.5% (= exp(2.510/100)) increased risk of RRT, with lower and upper 95% confidence interval limits of 2.4% (= exp(2.330/100)) and 2.7% (= exp(2.691/100)), respectively. Baseline age was also associated with the hazard for RRT, but unexpectedly in the negative direction; the corresponding estimated hazard ratio per year of age was 0.971 (95% CI 0.963, 0.979). Patients whose base hospital was not SRFT were estimated to have 44.5% higher hazard for RRT than patients whose base hospital was SRFT. Patients who were ex or current smokers were estimated to have a 43.6% higher hazard for RRT than patients who never smoked. On the other hand, there were no differences between males and females, between alcohol consumers and abstainers from alcohol, between diabetic and non-diabetic patients or between patients who did or did not have at least one comorbidity event; related *p*-values were 0.267, 0.647, 0.409 and 0.815, respectively.

Table 3. Estimated parameters and related 95% confidence intervals (95% CI), standard errors (SE), *p*-values (p), hazard ratios (HR) and related 95% confidence intervals, in analysis of CRISIS data set with Cox model with time-varying covariate for RRT as the event

Variable	Estimate (95% CI)	SE	*p*	HR (95% CI)
log(eGFR)	−2.510 (−2.691, −2.330)	0.092	<0.001	0.081 (0.068, 0.097)
Baseline age (years)	−0.030 (−0.038, −0.021)	0.004	<0.001	0.971 (0.963, 0.979)
Hospital base	−0.368 (−0.645, −0.091)	0.141	0.009	0.692 (0.525, 0.913)
Gender	0.139 (−0.106, 0.385)	0.125	0.267	1.149 (0.899, 1.469)
Smoking	0.362 (0.112, 0.612)	0.128	0.005	1.436 (1.119, 1.844)
Alcohol	0.055 (−0.180, 0.290)	0.120	0.647	1.056 (0.835, 1.336)
Diabetes	−0.109 (−0.366, 0.149)	0.131	0.409	0.897 (0.693, 1.161)
Comorbidity	−0.043 (−0.403, 0.317)	0.184	0.815	0.958 (0.669, 1.373)

Joint model

The longitudinal sub-model of the joint model was a random-intercept-and-random-slope model, as given in Equation 7. Results are shown in Table 4. The results for the longitudinal sub-model were consistent with the results from the separate longitudinal analysis. This might be explained by the fact that the modelling assumptions are the same and the censoring of the eGFR measurements due to RRT is not severe. The differences in magnitudes of the parameter estimates were negligible and there was no material difference in terms of statistical significance. In contrast, material differences were found in the results for the survival processes. This shows the importance of recognizing the measurement error in the observed values of eGFR. Hence, in what follows we discuss only the results for the survival sub-models.

Most importantly, we found that a 1% reduction in GFR level was associated with a 3.7% (95% CI 3.3%, 4.1%)

Table 4. Results for the joint modelling analysis of CRISIS data set. For the longitudinal sub-model, estimated parameters and related 95% confidence intervals (95% CI), standard errors (SE), *p*-values (p) and percentage relative effects (RE %) are reported. For the survival sub-model, estimated parameters, related 95% confidence intervals, standard errors, *p*-values, hazard ratios (HR) and related 95% confidence intervals are reported

Variable	Estimate (95% CI)	SE	*p*	RE%
Longitudinal sub-model				
Intercept	3.614 (3.495, 3.732)	0.060	<0.001	NA
Baseline age (years)	−0.005 (−0.007, −0.003)	0.001	<0.001	−0.5
Follow-up	−0.073 (−0.081, −0.064)	0.004	<0.001	−7.0
Hospital base	0.133 (0.081, 0.185)	0.027	<0.001	14.2
Gender	0.077 (0.026, 0.129)	0.026	0.003	8.0
Smoking	0.021 (−0.032, 0.074)	0.027	0.430	2.2
Alcohol	0.038 (−0.011, 0.088)	0.025	0.130	3.9
Diabetes	−0.096 (−0.149, −0.044)	0.027	<0.001	−9.2
Comorbidity	−0.029 (−0.093, 0.035)	0.032	0.373	−2.8
Variable	**Estimate (95% CI)**	**SE**	***p***	**HR (95% CI)**
Survival sub-model				
log(GFR)	−3.656 (−4.042, −3.270)	0.197	<0.001	0.026 (0.018, 0.038)
Baseline age (years)	−0.036 (−0.046, −0.026)	0.005	<0.001	0.964 (0.955, 0.974)
Hospital base	−0.190 (−0.509, 0.128)	0.162	0.241	0.827 (0.601, 1.136)
Gender	0.204 (−0.085, 0.493)	0.148	0.167	1.226 (0.918, 1.637)
Smoking	0.480 (0.186, 0.774)	0.150	0.001	1.616 (1.204, 2.169)
Alcohol	0.036 (−0.239, 0.311)	0.140	0.796	1.037 (0.788, 1.365)
Diabetes	0.085 (−0.210, 0.380)	0.151	0.572	1.089 (0.811, 1.463)
Comorbidity	−0.112 (−0.534, 0.310)	0.215	0.603	0.894 (0.586, 1.363)

increased hazard for RRT, compared with 2.7% (95% CI 2.4%, 2.9%) previously obtained in the separate analysis of the survival outcome. There was no difference between patients whose base hospital was or was not SRFT (*p*-value = 0.552), whereas in the separate analysis of survival outcome, base hospital was associated with the hazard for RRT (*p*-value = 0.025).

Discussion

Our aim in this paper has been to present an introductory tutorial on simultaneous analysis of repeated measurement and time-to-event outcome data, using the *joint modelling* approach that features most prominently in the biostatistical literature. This approach incorporates the most widely used methods that have been developed for separate analysis of the two types of outcome, namely linear mixed effects modelling and Cox proportional hazards modelling with frailty, and combines the two by linking their respective random effects. The principal advantage of this approach over separate analyses of each outcome is the correct treatment of noisy and incompletely observed time-varying covariate information, which enables unbiased estimation of the relationship between the two.

We have reported an analysis of a data set from the CRISIS cohort on CKD patients, where the repeated measurements are serial eGFR measurements and the survival outcome is time to initiation of RRT. The results demonstrate the usefulness of kidney function as a predictor for the hazard for initiation of RRT. This was substantially underestimated in a separate analysis of time to initiation of RRT treating eGFR as a time-varying covariate, because the separate analysis fails to take account of the measurement error in eGFR. In general, measurement error in a covariate biases the estimate of the associated regression parameter towards zero.[5,10,28,29]

We found an unexpected result that increased baseline age was associated with a decreased hazard for initiation of RRT. This is most likely explained by the fact that regression associations do not generally equate to causal relationships. Another explanation might be that younger patients with poor kidney function were given priority for RRT.

Our analysis of time to initiation of RRT as the survival outcome alone, treating death as a right-censored event time, could be criticised for failing to take account of the status of RRT and death as asymmetrical competing risks, in the sense that RRT necessarily precedes death, whereas death automatically censors initiation of RRT. Joint models for such competing risks have been considered, for example in the work of Williamson *et al.*,[30] but are beyond the scope of this paper; see also Chapter 5.5 of Rizopoulos (2012).[8] Similarly, we did not include proteinurea or blood pressure data in our analyses because both are subject to measurement error and are collected intermittently at irregular times. We would argue that analysis of proteinurea and blood pressure together with eGFR strictly requires *multivariate* joint modelling. This is an area of current

methodological research and cannot be implemented routinely in freely available software packages.

The joint modelling framework we considered in this paper is also called the shared random effects model. An alternative approach to this formulation is the latent class model. For a recent review, see Proust-Lima *et al.* (2014).[31] Latent class models can be fitted by the R package lcmm.[32]

We have presented joint modelling results obtained by ML estimation. An alternative approach for parameter estimation is Bayesian inference.[33,34] This can be applied to a limited class of models with the R package JMbayes.[35] Model fit can be assessed by Akaike or Bayesian Information Criteria.[36] The sensitivity of the fixed effects parameters in the longitudinal sub-model might be investigated by, for example, index of local sensitivity to non-ignorability.[37] Diagnostic tools can be applied to multiple-imputation based empirical residuals for the longitudinal sub-model of the joint models.[38] Methods to inspect the association between longitudinal and time-to-event data, to investigate empirical residuals of the longitudinal sub-model after fitting a joint model and to detect influential observations within the joint modelling framework can be found in Dobson and Henderson (2003).[39]

Software for the separate analysis of longitudinal and survival data is widely available. For the results reported in this paper, we used the R packages nlme[40] and survival[41] for longitudinal and survival data analysis, respectively. Software for joint modelling is becoming increasingly available in statistical packages, for example in the R packages JM,[26] joineR[27] and JMbayes.[35] The R scripts for the analyses reported in this paper are provided as Supplementary data, available at *IJE* online.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

This work was supported by Lancaster University through an e-Health Research Centre Studentship for Özgür Asar.

Conflict of interest: None declared.

References

1. Diggle PJ, Heagerty P, Liang K-Y *et al*. *Analysis of Longitudinal Data*. 2002, 2nd edn. Oxford. UK: Oxford University Press.
2. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. 2011, 2nd edn. Hoboken, NJ: John Wiley & Sons.
3. Kalbfeisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. 2002, 2nd edn. New York, NY: John Wiley & Sons.
4. Kleinbaum DG, Klein M. *Survival Analysis: A Self Learning Text*. 2012, 3rd edn. New York, NY: Springer.
5. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997;**53**: 330–39.
6. Henderson R, Diggle PJ, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000;**4**: 465–80.
7. Diggle PJ, Sousa I, Chetwynd A. Joint modelling of repeated measurements and time-to-event outcomes: the fourth Armitage lecture. *Stat Med* 2007;**26**:2981–98.
8. Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data*. 2012, Boca Raton, FL: Chapman & Hall/CRC.
9. Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *J Am Stat Assoc* 1995;**90**:27–37.
10. Ibrahim JG, Chu H, Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. *J Clin Oncol* 2010;**28**:2796–801.
11. Andrinopoulou E-R, Rizopoulos D, Jin R *et al*. An introduction to mixed models and joint modeling: analysis of valve function over time. *Ann Thorac Surg* 2012;**93**:1765–72.
12. Garre FG, Zwinderman AH, Geskus RB *et al*. A joint latent class changepoint model to improve the prediction of time to graft failure. *J R Stat Soc A Stat Soc* 2008;**171**:299–308.
13. Abdi ZD, Essig M, Rizopoulos D *et al*. Impact of longitudinal exposure to mycophenolic acid on acute rejection in renal-transplant recipients using a joint modeling approach. *Pharmacol Res* 2013;**72**:52–60.
14. R Development Core Team. *R: A Language and Environment for Statistical Computing*. 2013, Vienna: R Foundation for Statistical Computing.
15. Hoefield RA, Kalra PA, Baker P *et al*. Factors associated with kidney disease progression and mortality in a referred CKD population. *Am J Kidney Dis* 2010;**56**:1072–81.
16. Eddington H, Hoefield R, Sinha S *et al*. Serum phosphate and mortality in patients with chronic kidney disease. *Clin J Am Soc Nephrol* 2010;**5**:2251–57.
17. Levey AS, Bosch JP, Lewis JB *et al*. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med* 1999;**130**:461–70.
18. Carroll RJ, Ruppert D, Stefanski LA *et al*. *Measurement Error in Nonlinear Models: A Modern Perspective*, 2006, 2nd edn. Boca Raton, FL: Chapman & Hall/CRC.
19. Little RJA, Rubin DB. *Statistical Analysis With Missing Data*. 2002, New York, NY: John Wiley & Sons.
20. Molenberghs G, Kenward MG. *Missing Data in Clinical Studies*. 2007, New York, NY: John Wiley & Sons.
21. Ibrahim JG, Molenberghs G. Missing data methods in longitudinal studies: a review. *Test* 2009;**18**:1–43.
22. Laird NM, Ware JH. Random effects models for longitudinal data. *Biometrics* 1982;**38**:963–74.
23. Pawitan Y. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. 2001, Oxford, UK: Oxford University Press.
24. Cox DR. Regression models and life tables. *J R Stat Soc B Stat Methodol* 1972;**34**:187–220.
25. Cox DR. Partial likelihood. *Biometrika* 1975;**62**:269–76.

26. Rizopoulos D. JM: an R package for the joint modelling of longitudinal and time-to-event data. *J Stat Soft* 2010;**35**: 1–33.
27. Philipson P, Sousa I, Diggle PJ *et al*. 2012. joineR: joint modelling of repeated measurements and time-to-event data, R package version 1.0-3. http://CRAN.R-project.org/package=joineR (5 May 2014, date last accessed).
28. Prentice RL. Covariate measurement error and parameter estimation in a failure time regression model. *Biometrika* 1982;**69**: 331–42.
29. Sweeting M, Thompson SG. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometr J* 2011;**53**: 750–63.
30. Williamson P, Kolamunnage-Dona R, Philipson P *et al*. Joint modelling of longitudinal and competing risks data. *Stat Med* 2008;**27**:6426–38.
31. Proust-Lima C, Séne M, Taylor J *et al*. Joint latent class models for longitudinal and time-to-event data: a review. *Stat Methods Med Res* 2014;**23**:74–90.
32. Proust-Lima C, Philipps V, Diakite A *et al*. 2014. lcmm: Estimation of extended mixed models using latent classes and latent processes, R package version 1.6.4. http://CRAN.R-project.org/package=lcmm (26 September 2014, date last accessed).
33. Ibrahim J, Chen M, Sinha D. *Bayesian Survival Analysis* 2001. New York, NY: Springer.
34. Guo X, Carlin B. Separate and joint modeling of longitudinal and event time data using standard computer packages. *Am Stat* 2004;**58**:16–24.
35. Rizopoulos D. 2014. JMbayes: joint modeling of longitudinal and time-to-event data under a Bayesian approach. R package version 0.6-1. http://CRAN.R-project.org/package=JMbayes (5 May 2014, date last accessed).
36. Zhang D, Chen M-H, Ibrahim JG *et al*. Assessing model fit in joint models of longitudinal and survival data with applications to cancer clinical trials. *Stat Med*;**33**:4715–33.
37. Viviani S, Rizopoulos D, Alfó M. Local sensitivity to non-ignorability in joint models. *Stat Model* 2014;**14**:205–28.
38. Rizopoulos D, Verbeke G, Molenberghs G. Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics* 2010;**66**:20–29.
39. Dobson A, Henderson R. Diagnostics for joint longitudinal and dropout time modelling. *Biometrics* 2003;**59**:741–51.
40. Pinheiro J, Bates D, DebRoy S *et al*. 2013. nlme: linear and nonlinear mixed effects models, R package version 3.1-109. http://CRAN.R-project.org/package=nlme (18 November 2014, date last accessed).
41. Therneau T. 2013. A package for survival analysis in S. R package version 2.37-4. http://CRAN.R-project.org/package=survival (18 November 2014, date last accessed).