BMC
Medical Research Methodology

RESEARCH ARTICLE   **Open Access**

Adjusting for measurement error in baseline prognostic biomarkers included in a time-to-event analysis: a joint modelling approach

Michael J Crowther[1]*, Paul C Lambert[1,2] and Keith R Abrams[1]

Abstract

Background: Methodological development of joint models of longitudinal and survival data has been rapid in recent years; however, their full potential in applied settings are yet to be fully explored. We describe a novel use of a specific association structure, linking the two component models through the subject specific intercept, and thus extend joint models to account for measurement error in a biomarker, even when only the baseline value of the biomarker is of interest. This is a common occurrence in registry data sources, where often repeated measurements exist but are simply ignored.

Methods: The proposed specification is evaluated through simulation and applied to data from the General Practice Research Database, investigating the association between baseline Systolic Blood Pressure (SBP) and the time-to-stroke in a cohort of obese patients with type 2 diabetes mellitus.

Results: By directly modelling the longitudinal component we reduce bias in the hazard ratio for the effect of baseline SBP on the time-to-stroke, showing the large potential to improve on previous prognostic models which use only observed baseline biomarker values.

Conclusions: The joint modelling of longitudinal and survival data is a valid approach to account for measurement error in the analysis of a repeatedly measured biomarker and a time-to-event. User friendly Stata software is provided.

Background

Many biomarkers such as systolic blood pressure (SBP) have been identified as key prognostic factors in the development and validation of cardiovascular risk scores [1,2]. However, often only baseline values of these biomarkers are used, despite the existence of repeated measures, especially in registry sources such as the General Practice Research Database (GPRD) [3]. Furthermore, biomarkers are often measured with error. Failing to adjust for such measurement error leads to estimates being biased towards the null [4].

A joint model of longitudinal and survival data allows us to investigate the relationship between a repeatedly measured biomarker, subject to measurement error, such as SBP, and the time to an event of interest, such as time to non-fatal stroke. The approach which has dominated the methodological literature involves linking the two component submodels using shared random effects [5,6]. From a classical perspective, these methods require computationally intensive numerical integration, which is difficult to implement. However, due to the recent introduction of user-friendly software in R [7,8] and Stata [9], these models are starting to find their place in applied research [10,11], but the potential uses of and forms of the association parameters, linking the longitudinal and survival components, are yet to be fully explored. Alternatively, many authors have proposed a Bayesian approach,

*Correspondence: michael.crowther@le.ac.uk
[1]University of Leicester, Department of Health Sciences, Adrian Building, University Road, Leicester LE1 7RH, UK
Full list of author information is available at the end of the article

proving readily available BUGS code to implement the models [12,13].

The most commonly used association structures include the current value parameterisation [5]; whereby we directly link the value of the biomarker, as estimated by the longitudinal submodel, to survival, and the first derivative or slope [10]; allowing the investigation of the effect that the rate of change of the biomarker has on survival.

There is often interest in predicting prognosis based on an initial baseline measurement [1,2]. In this paper we investigate the use of the joint model framework with a random intercept association structure as an approach to adjust for measurement error, inherent in biomarkers such as SBP. By incorporating the repeated measures we thus make the most efficient use of the data available. In particular, as a prognostic model for future patients, we describe how this framework can be used to predict survival for new patients who will only have baseline measurements.

Methods

A joint model of longitudinal and survival data consists of two component submodels: the longitudinal submodel and the survival submodel. We define a set of baseline covariates, U_i, which can potentially differ between submodels. The longitudinal submodel allows us to model the trajectory of a repeatedly measured biomarker over time, adjusting for baseline covariates. The standard approach assumes a linear mixed effects model [14]. We observe

$$Y_i(t_{ij}) = W_i(t_{ij}) + \epsilon_{ij}, \qquad \epsilon_{ij} \sim \mathrm{N}(0, \sigma_e^2) \qquad (1)$$

with

$$W_i(t_{ij}) = X_i'(t_{ij})\boldsymbol{\beta} + Z_i'(t_{ij})\boldsymbol{b}_i + \boldsymbol{u}_i\boldsymbol{\delta} \qquad (2)$$

where $Y_i(t_{ij})$ is the observed longitudinal response for the ith patient measured at the jth time point. $W_i(t_{ij})$ is our true unobserved trajectory function consisting of design matrices $X_i'(t_{ij})$ and $Z_i'(t_{ij})$ for the fixed and random effects, $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$, respectively, where $\boldsymbol{b}_i \sim \mathrm{MVN}(0, \boldsymbol{\Sigma})$. We can incorporate flexibility here by allowing both $X_i'(t_{ij})$ and $Z_i'(t_{ij})$ to contain restricted cubic spline functions of measurement time [15]. We also have a vector of baseline covariates $\boldsymbol{u}_i \in \boldsymbol{U}_i$, and corresponding regression coefficients, $\boldsymbol{\delta}$. Finally, ϵ_{ij} is our normally distributed measurement error with constant variance σ_e^2. We further assume that the random effects and error term are independent, and that $\mathrm{cov}(\epsilon_{ij}, \epsilon_{ik}) = 0$ (where $j \neq k$).

The time-to-event submodel usually takes the form of a proportional hazards model

$$h_i(t) = h_0(t)\exp(\alpha_1 W_i(t) + \boldsymbol{\phi}\boldsymbol{v}_i) \qquad (3)$$

with $h_0(t)$ the baseline hazard function and $\boldsymbol{v}_i \in \boldsymbol{U}_i$ is a vector of baseline covariates with corresponding log hazard ratios, $\boldsymbol{\phi}$. The parameter α_1 is commonly named the association parameter, indicating the strength of association between the longitudinal biomarker and the time to event. If $\alpha_1 = 0$, then the joint model reduces to the two separate models and fitting a joint model will not prove advantageous. This parameterisation assumes the hazard is dependent on the biomarker through its current value. This form of association is one of many ways to link the two component sub-models. The baseline hazard function, $h_0(t)$, can be modelled using a parametric distribution, most frequently the Weibull, or less restrictively using flexible parametric survival models [16], or of course can be left unspecified [17]. However, an unspecified baseline hazard function leads to underestimation of the standard errors of parameter estimates [18], and consequently bootstrapping is required to obtain appropriate standard errors.

For illustration, we let $W_i(t_{ij})$, the longitudinal submodel, be a linear function of time where the intercept and slope varies between subjects

$$W_i(t_{ij}) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} \qquad (4)$$

giving a model with a random intercept and random linear slope. As an alternative way of linking the component models to that of Equation (3), we may link elements of the trajectory function, $W_i(t_{ij})$, to the hazard directly. For example, we can link the subject specific baseline biomarker values through the intercept association structure, where

$$h_i(t) = h_0(t)\exp\left[\alpha_2(\beta_0 + b_{0i}) + \boldsymbol{\phi}\boldsymbol{v}_i\right] \qquad (5)$$

in this expression α_2 now estimates the strength of the association between the patient specific baseline biomarker values, as estimated by the longitudinal submodel, and the time-to-event. This way we can let the risk of event depend directly on the subject specific value of the biomarker at time $t = 0$.

If interest lies in prediction when a new patient is observed at baseline, the issue of measurement error can be accounted for through this approach. A benefit of this association structure also lies in the evaluation of the joint likelihood. Under most parametric survival submodels (e.g. Weibull distribution) and time-dependent association structures (eg. current value), numerical quadrature is required to integrate out not only the random effects, but under Equation (3), nested quadrature is also required to evaluate the cumulative hazard function. Under the time-independent association structure of Equation (5), we avoid this nested quadrature as the cumulative hazard function has an analytically tractable form, which provides computational benefits.

As discussed in the introduction, this model formulation can be an alternative to the standard approach of using the observed baseline biomarker value

$$h_i(t) = h_{0i}(t) \exp(\alpha_3 Y_{0i} + \boldsymbol{\phi} \boldsymbol{v}_i) \tag{6}$$

where Y_{0i} is the observed baseline biomarker value and α_3 is the log hazard ratio for a one unit increase in the observed baseline biomarker value. Although simple to fit, Equation (6) does not account for potential measurement error in Y_{0i}.

Simulation study
In order to assess the performance of the standard approach of including observed biomarker values, compared to the full joint model described above, we evaluated both through simulation [19]. For ease of exposition we assume a longitudinal model with random intercept and slope, assuming a continuous biomarker of interest with

$$W_i(t_{ij}) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij}$$

where $\beta_0 = \beta_1 = 0$, and $b_{0i} \sim \mathrm{N}(0, 1)$, $b_{1i} \sim \mathrm{N}(0, 0.25^2)$ with correlation between (b_{0i}, b_{1i}) of 0.25. Observed measurements are then generated from $Y_{ij} \sim \mathrm{N}(W_i(t_{ij}), \sigma_e^2)$, where t_{ij} is the time of the jth measurement for the ith patient. We vary σ_e from $\{0.1, 0.5, 1\}$.

We assume a Weibull baseline hazard function with $\lambda = 0.1$ and $\gamma = 1.5$. A binary variable, X_1 to represent treatment group was generated from $\mathrm{Bin}(1, 0.5)$, with an associated log hazard ratio of $\phi_1 = -0.5$. A continuous covariate, X_2, to represent age at baseline was generated from $\mathrm{N}(65, 12)$ with an associated log hazard ratio of $\phi_2 = 0.01$. We then generate survival times from a Weibull distribution where the hazard is defined as $h(t) = h_0(t) \exp(\alpha_2 \beta_{0i} + \phi_1 X_1 + \phi_2 X_2)$, with α_2 the association parameter, indicating the effect of a one unit increase in the value of the subject specific intercept on the risk of event. We vary $\alpha_2 = \{-0.5, 0.25, 0.25, 0.5\}$. Each simulation contained 300 patients with up to 5 annual measurements (including baseline), and administrative censoring at 5 years. This corresponds to an approximate 18.9% survival proportion at 5 years (calculated at the mean of covariate values, $X_1 = \frac{1}{2}, X_2 = 65$ and $\beta_{0i} = 0$). To each dataset we fit a Weibull proportional hazards model including the observed baseline measurement, and a Weibull-based joint model with the random intercept association structure. We adjust for age and treatment in the survival submodel. Each scenario is simulated 1000 times.

To illustrate the varying measurement error standard deviations used in the simulation scenarios, we show in Figure 1 observed longitudinal measurements from the same 100 patients with $\sigma_e = \{0.1, 0.5, 1\}$, and when

$\alpha = 0.25$. Figure 1 illustrates that as the measurement error standard deviation increases, the variability in the observed biomarker values increases.

The GPRD cohort
The General Practice Research Database (GPRD) Group has obtained ethical approval from a Multi-centre Research Ethics Committee (MREC) for all purely observational research using GPRD data; namely, studies which do not include patient involvement The core work of the GPRD is covered by MREC approval granted by the Trent Multi- Research Ethics Committee (REC reference number 05/MRE04/87) and this study was approved by the GPRD Independent Scientific Advisory Committee (ISAC) (Protocol number 09_094). This study is based in part on data from GPRD obtained under licence from the UK Medicines and Healthcare Product Regulatory Agency (MHRA). However, the interpretation and conclusions contained in this study are those of the authors alone.

The example cohort used to illustrate the methods in this paper consists of 4,850 obese patients diagnosed with type 2 diabetes mellitus. We have 107,347 measurements of SBP, with maximum follow-up of 22 years. There were 278 stroke events observed.

In Figure 2 we show the observed SBP measurements for 9 randomly selected patients, who had at least 10 measurements, illustrating some nonlinear trajectories. To accommodate such nonlinearities we can use restricted cubic splines in the linear mixed effects submodel. In particular, we specify the following longitudinal submodel

$$\begin{aligned} W_i(t_{ij}) = {} & (\beta_0 + b_{0i}) + \beta_1 \mathrm{age}_i + \beta_2 \mathrm{sex}_i + \beta_3 \mathrm{BMI}_i \\ & + (\boldsymbol{\beta}_F s_F(t_{ij}; \mathbf{k}_F) + \mathbf{b}_R s_R(t_{ij}; \mathbf{k}_R)) \end{aligned} \tag{7}$$

Where $s_F(t_{ij}; \mathbf{k}_F)$ is the restricted cubic spline basis of measurement time with corresponding fixed effects, $\boldsymbol{\beta}_F$, with knot locations $\mathbf{k}_F$, and $s_R(t_{ij}; \mathbf{k}_R)$ is the restricted cubic spline basis of measurement time with corresponding random effects, $\mathbf{b}_R$, and knot locations $\mathbf{k}_R$.

Preliminary modelling of the longitudinal data can be conducted to guide model selection, in particular, the degrees of freedom for the spline terms capturing the underlying longitudinal trajectory over time.

To allow flexibility in the survival submodel we use the flexible parametric survival model [16,20], which models the baseline log cumulative hazard function using restricted cubic splines. We can once again undertake seperate analysis of just the survival data to inform model selection. In particular, we can use the AIC and BIC to guide the selection of the number of degrees of freedom to capture the baseline hazard function, following
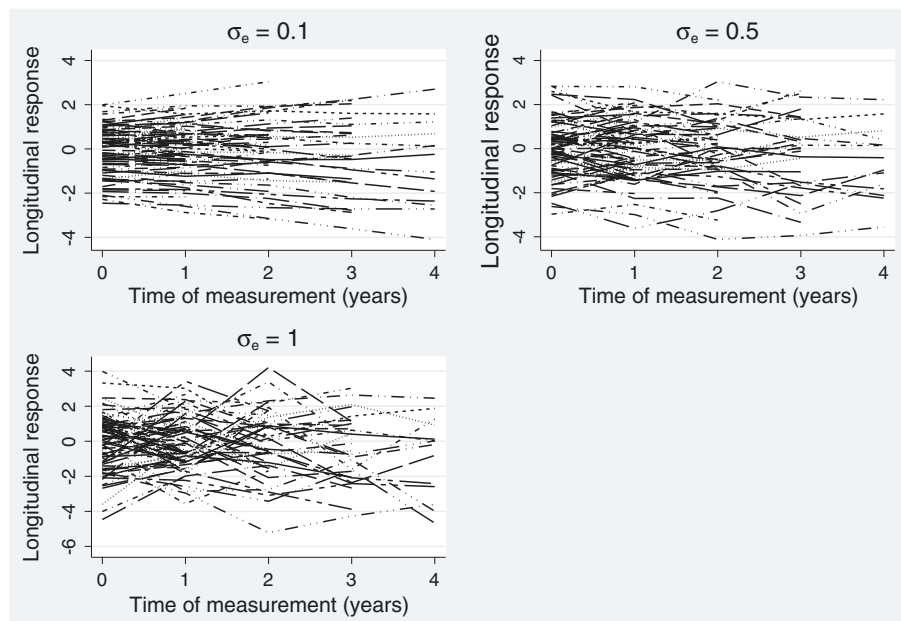
Figure 1 Example simulated observed longitudinal measurements with varying measurement error standard deviation.
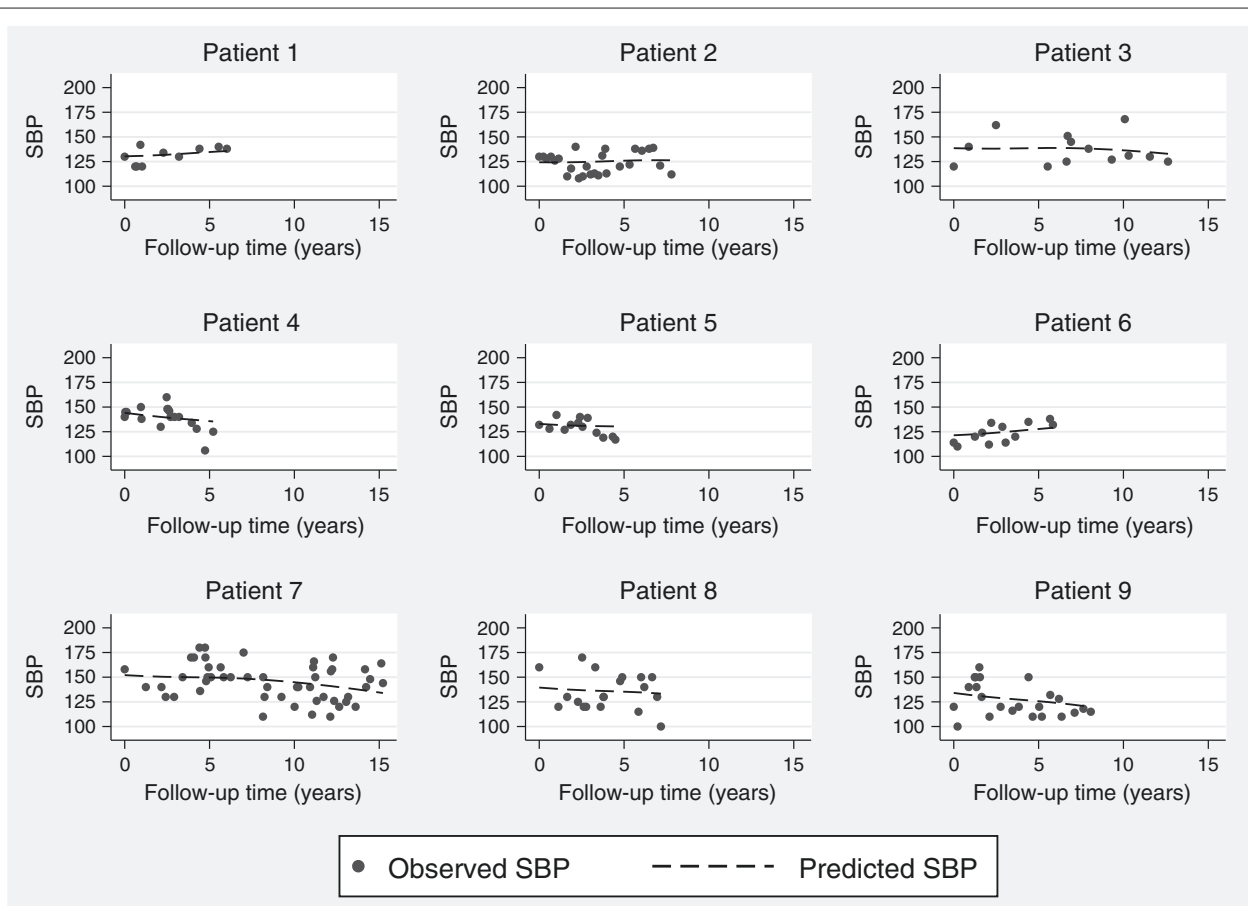

Figure 2 Longitudinal response measurements for SBP for 9 randomly selected patients who had at least 10 measurements. The dashed line represents the fitted longitudinal trajectories based on the joint model.

Rutherford et al. (2013) [21]. Our final joint model is then

$$\log[H_i(t)] = \log[H_0(t)] + \phi_1 \text{age}_i + \phi_2 \text{sex}_i + \phi_3 \text{BMI}_i + \alpha_2(\beta_0 + b_{0i})$$

(8)

where

$$\log[H_0(t)] = s(\log(t); \gamma, \mathbf{k}_S)$$ (9)

where the baseline log cumulative hazard function, $\log[H_0(t)]$, is expanded into a restricted cubic spline function of $\log(t)$, $s(\log(t); \gamma, \mathbf{k}_S)$, with knot locations $\mathbf{k}_S$ and coefficient vector, γ. This framework has recently been incorporated into a joint model [22]. In each submodel we adjust for the baseline effects of age, sex and BMI. We fit the joint model with the random intercept association structure shown in Equation (5). For comparison, we also apply the standard flexible parametric survival model, adjusting for observed baseline SBP, age, sex and BMI.

Results

Simulation study results

Bias and coverage estimates for the association parameter are presented in Table 1. Under the standard Weibull model, we observe increasing bias in the estimates of the association between baseline biomarker values and survival, as the magnitude of the measurement error standard deviation, σ_e, increases. In parallel we observe very poor coverage probabilities under the Weibull approach. For example, with $\alpha = 0.5$ and $\sigma_e = 1$, we observe

bias of -0.261 (percentage bias of -52.2%) and coverage of 0.4%. In contrast, under the joint modelling approach we observe minimal bias and coverage probabilities close to 95% across all scenarios.

Analysis of GPRD cohort

We now present the analysis of the GPRD cohort. In all analyses we use SBP/10 so that a unit increase in SBP/10 represents a clinically meaningful 10 unit increase in SBP. Our primary interest is the association between baseline SBP and the risk of stroke. Baseline ($t_{ij} = 0$) corresponds to when each patient entered the cohort, i.e. the time of first SBP measurement.

We began by assuming a random intercept and selecting the degrees of freedom for the fixed spline terms using the AIC and BIC. In this case, both selected five degrees of freedom for $s_F(t_{ij}; \mathbf{k}_F)$, with an AIC of 417565.8 and BIC of 417604.1. For the random splines of time we assumed a linear term, which equates to one spline term for $s_R(t_{ij}; \mathbf{k}_R)$. This allows a very flexible form to take into account the variation in SBP over time. We further adjust for age, sex and Body-Mass Index (BMI) at baseline.

For the flexible parametric survival submodel, both AIC and BIC selected two degrees of freedom, with an AIC of 2408.7173 and BIC of 2430.483. If one degree of freedom had been selected, then this would be equivalent to a Weibull survival model.

Results are presented in Table 2. Under the standard flexible parametric survival model we observe a hazard ratio for a ten unit increase in baseline SBP of 1.111 (95% CI: 1.051, 1.172). Under a joint model we observe an increased hazard ratio of 1.198 (95% CI: 1.107, 1.298). The

Table 1 Simulation results of the association parameter, α

True α	True σ_e	Weibull				Joint model			
		Bias	% bias	MSE	CP	Bias	% bias	MSE	CP
	0.1	-0.001	-0.2	0.006	94.8	0.005	0.9	0.006	95.3
0.50	0.5	-0.105	-21.1	0.016	65.4	0.005	0.9	0.007	95.6
	1.0	-0.261	-52.1	0.071	0.4	0.008	1.6	0.012	94.8
	0.1	0.002	1.0	0.005	94.4	0.005	2.0	0.006	94.3
0.25	0.5	-0.046	-18.5	0.007	89.0	0.007	2.7	0.007	94.5
	1.0	-0.123	-49.2	0.018	34.1	0.010	4.1	0.009	94.8
	0.1	0.003	-1.3	0.006	93.8	0.001	-0.2	0.006	94.0
-0.25	0.5	0.051	-20.6	0.007	87.1	0.000	-0.1	0.007	94.2
	1.0	0.127	-50.7	0.019	29.7	-0.002	0.9	0.009	94.6
	0.1	0.000	-0.1	0.006	96.6	-0.005	1.0	0.006	95.9
-0.50	0.5	0.104	-20.9	0.015	66.7	-0.006	1.1	0.007	95.7
	1.0	0.260	-52.0	0.070	0.4	-0.010	2.0	0.012	94.5

MSE - mean square error.
CP - coverage probability.
σ_e - standard deviation of the measurement error.

Table 2 Results from applying a flexible parametric proportional hazards model adjusting for observed baseline systolic blood pressure, and a full joint model using the intercept association structure

	Standard FPSM			Joint model		
	Coefficient	95% CI		Coefficient	95% CI	
Survival model:						
Baseline SBP/10 (α_2)	0.105	0.050	0.159	0.181	0.102	0.261
Age (years)	0.048	0.036	0.060	0.050	0.038	0.062
Sex (male)	0.011	-0.233	0.254	-0.010	-0.253	0.234
BMI (kg/m^2)	0.011	-0.015	0.037	0.013	-0.012	0.039
Longitudinal model:						
Intercept	-	-	-	13.006	12.629	13.382
Age (years)	-	-	-	0.025	0.022	0.029
Sex (male)	-	-	-	-0.252	-0.332	-0.171
BMI (kg/m^2)	-	-	-	0.003	-0.005	0.011
RCS1	-	-	-	-0.080	-0.121	-0.039
RCS2	-	-	-	-0.006	-0.019	0.006
RCS3	-	-	-	-0.001	-0.010	0.007
RCS4	-	-	-	0.003	0.000	0.006
RCS5	-	-	-	0.000	-0.001	0.001
σ_e	-	-	-	1.522	1.515	1.528

FPSM - Flexible Parametric Survival Model.
RCS - Restricted Cubic Spline.

increased effect using a joint model is consistent with that observed in the simulation study, i.e. that the bias in the standard survival model is towards the null. The fitted trajectories seen in Figure 2 appear to capture the subject-specific measurements well, although some panels appear to only require a linear trend.

We illustrate how the bias from the standard approach increases with SBP in Figure 3, showing predictions from both models for a female patient aged 60, with low (90), medium (130) and high (200) SBP baseline measurements. To quantify the differences, at 10 years under the standard model we observe a survival probability of 0.881 for a SBP of 200, compared to 0.816 under the full joint model.

Discussion

A wealth of patient data is becoming available in registry sources such as the GPRD, providing extensive opportunities to utilise the joint modelling framework. We have shown that by incorporating repeated measures of a biomarker within a unified joint model framework, we reduce bias due to measurement error, even when only the baseline level of the biomarker is predictive of survival. As illustrated in the simulation study, ignoring measurement error in biomarkers such as blood pressure can lead to a marked underestimation of covariate effects. In our application, through the use of restricted

cubic splines in the linear mixed effects submodel, we can model highly nonlinear trajectories over time, compared to linear slope models. Furthermore, the flexible parametric survival submodel can also capture complex baseline hazard functions, an important component when predicting survival at the patient level [22].

Given that, to our knowledge, all current cardiovascular risk scores only use baseline measures, with no
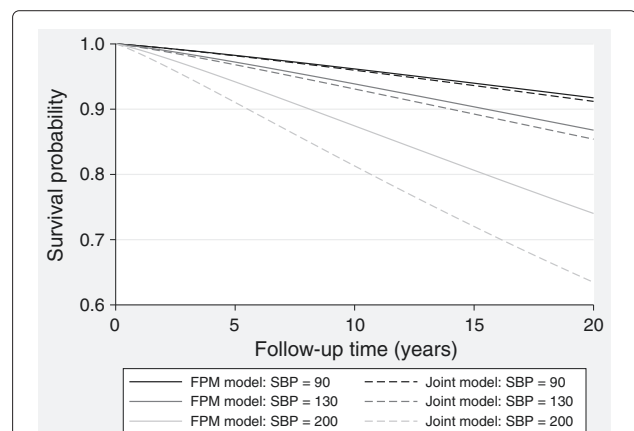
Figure 3 Predicted survival from the flexible parametric survival model and joint model, for a female, aged 60 years, BMI of 30, with SBP of 90, 130 or 200.

adjustment for measurement error, the prospects of utilising this framework to improve prognostic risk scores is quite substantial. Predicting survival for a new patient using this framework follows naturally, as often only a first baseline biomarker observation will be available. However, such a modelling approach also allows a dynamic risk prediction approach to be adopted, whereby a patient's estimated future risk is updated as each new biomarker value is obtained [23]. Such an approach could enable response to treatment to be monitored and patients counselled accordingly.

In the analysis of the GPRD cohort, we incorporated flexibility in both the longitudinal submodel through the use of restricted cubic splines, and the flexible parametric survival submodel. Given that both submodels require choosing the number of degrees of freedom, a simple sensitivity analysis can be undertaken to assess knot locations and number of knots. We showed recently that the flexible parametric survival submodel is very robust to both knot placement and number of knots within a joint model framework [22], and furthermore, an extensive simulation study has been conducted by Rutherford et al. (2013), which showed excellent performance of the flexible parametric model to capture simple and complex baseline hazard functions [21]. Furthermore, given that primary interest was in the survival component, and the estimate of association, often modelling the longitudinal component with a suitable sensible functional form will provide an improved estimate compared to simplistic approaches of seperate modelling.

In this paper we have concentrated on a specific association structure linking the 2 component submodels; however, it may be of interest to investigate linking multiple components of a biomarkers trajectory to the time to an event of interest. For example, recent work by Rothwell et al. (2010) [24] has shown associations between not only baseline blood pressure, but also variability over time as important predictors of cardiovascular events. Furthermore, we have only compared the standard approach of adjusting for observed baseline biomarker values to the full joint model. It would be of interest to compare alternative approaches for adjusting for measurement error, not only in baseline biomarkers, but also under a time-dependent association structure [25,26].

Extensions to the modelling framework include incorporating multiple biomarkers. In particular, in our example we modelled SBP over time, whilst adjusting for baseline BMI. It may be of interest to model not only SBP but also the inter-relationships between different biomarkers such as BMI, and how they are related to an event of interest [13].

To facilitate the use of the methods in practice, user friendly Stata software, written by the first author, is available, with a variety of survival model choices and association structures, including those discussed in this article [9,27]. To illustrate computational aspects of the framework, the presented joint model applied to the cohort took just over 13 minutes to converge on a standard laptop computer.

Conclusion
The joint modelling of longitudinal and survival data is a valid approach to account for measurement error in the analysis of a repeatedly measured biomarker and a time to event. User friendly Stata software is provided.

Author details
[1]University of Leicester, Department of Health Sciences, Adrian Building, University Road,Leicester LE1 7RH, UK. [2]Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Box 281, S-171 77 Stockholm, Sweden.

References
1. Conroy RM, Pyorala K, Fitzgerald AP, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetiere P, Jousilahti P, Keil U, Njolstad I, Oganov RG, Thomsen T, Tunstall-Pedoe H, Tverdal A, Wedel H, Whincup P, Wilhelmsen L, Graham IM, SCOREpg: **Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project.** *Eur Heart J* 2003, **24**(11):987–1003.
2. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P: **Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study.** *BMJ* 2007, **335**(7611):136. [http://dx.doi.org/10.1136/bmj.39261.471806.55]
3. Ara R, Blake L, Gray L, Hernandez M, Crowther M, Dunkley A, Warren F, Jackson R, Rees A, Stevenson M, Abrams K, Cooper N, Davies M, Khunti K, Sutton A: **What is the clinical effectiveness and cost-effectiveness of using drugs in treating obese patients in primary care? A systematic review.** *Health Technol Assess* 2012, **16**(5):1–202. [http://dx.doi.org/10.3310/hta16050]
4. Prentice RL: **Covariate measurement errors and parameter estimation in a failure time regression model.** *Biometrika* 1982, **69**(2):331–342. [http://www.jstor.org/stable/2335407]
5. Wulfsohn MS, Tsiatis AA: **A joint model for survival and longitudinal data measured with error.** *Biometrics* 1997, **53**:330–339.
6. Henderson R, Diggle P, Dobson A: **Joint modelling of longitudinal measurements and event time data.** *Biostatistics* 2000, **1**(4):465–480.

7. Rizopoulos D: **JM: An R package for the joint modelling of longitudinal and time-to-event data.** *J Stat Softw* 2010, **35**(9):1–33. [http://www.jstatsoft.org/v35/i09]

8. Philipson P, Sousa I, Diggle P, Williamson P, Kolamunnage-Dona R, Henderson R: **joineR - Joint modelling of repeated measurements and time-to-event data.** 2012. [http://cran.r-project.org/web/packages/joineR/index.html]

9. Crowther MJ, Abrams KR, Lambert PC: **Joint modeling of longitudinal and survival data.** *Stata J* 2013, **13**:165–184.

10. Wolbers M, Babiker A, Sabin C, Young J, Dorrucci M, Chêne G, Mussini C, Porter K, Bucher HC, CASCADE: **Pretreatment CD4 cell slope and progression to AIDS or death in HIV-infected patients initiating antiretroviral therapy–the CASCADE collaboration: a collaboration of 23 cohort studies.** *PLoS Med* 2010, **7**(2):e1000239. [http://dx.doi.org/10.1371/journal.pmed.1000239]

11. Ibrahim JG, Chu H, Chen LM: **Basic concepts and methods for joint models of longitudinal and survival data.** *J Clin Oncol* 2010, **28**(16):2796–2801.

12. Guo X, Carlin BP: **Separate and joint modeling of longitudinal and event time data using standard computer packages.** *Am Stat* 2004, **58**:16–24. [http://www.jstor.org/stable/27643494]

13. Rizopoulos D, Ghosh P: **A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event.** *Stat Med* 2011, **30**(12):1366–1380.

14. Laird NM, Ware JH: **Random-effects models for longitudinal data.** *Biometrics* 1982, **38**(4):963–974.

15. Durrleman S, Simon R: **Flexible regression models with cubic splines.** *Stat Med* 1989, **8**(5):551–561.

16. Royston P, Lambert PC: *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. College Station: Stata Press; 2011.

17. Cox DR: **Regression models and life-tables.** *J R Stat Soc Ser B Methodological* 1972, **34**(2):187–220.

18. Hsieh F, Tseng YK, Wang JL: **Joint modeling of survival and longitudinal data: likelihood approach revisited.** *Biometrics* 2006, **62**(4):1037–1043. [http://dx.doi.org/10.1111/j.1541-0420.2006.00570.x]

19. Burton A, Altman DG, Royston P, Holder RL: **The design of simulation studies in medical statistics.** *Stat Med* 2006, **25**(11):4279–4292.

20. Royston P, Parmar MKB: **Flexible parametric proportional hazards and proportional odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects.** *Stat Med* 2002, **21**(15):2175–2197.

21. Rutherford MJ, Crowther MJ, Lambert PC: **The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study.** *J Stat Comput Simul* 2013. [http://www.tandfonline.com/doi/abs/10.1080/00949655.2013.845890]

22. Crowther MJ, Abrams KR, Lambert PC: **Flexible parametric joint modelling of longitudinal and survival data.** *Stat Med* 2012, **31**(30):4456–4471. [http://dx.doi.org/10.1002/sim.5644]

23. Rizopoulos D: **Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data.** *Biometrics* 2011, **67**(3):819–829. [http://dx.doi.org/10.1111/j.1541-0420.2010.01546.x]

24. Rothwell PM, Howard SC, Dolan E, O'Brien E, Dobson JE, Dahlöf B, Sever PS, Poulter NR: **Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic hypertension.** *Lancet* 2010, **375**(9718):895–905. [http://dx.doi.org/10.1016/S0140-6736(10)60308-X]

25. Zucker DM: **A pseudo partial likelihood method for semiparametric survival regression with covariate errors.** *J Am Stat Assoc* 2005, **100**(472):1264–1277. [http://www.tandfonline.com/doi/abs/10.1198/016214505000000538]

26. Liao X, Zucker DM, Li Y, Spiegelman D: **Survival analysis with error-prone time-varying covariates: a risk set calibration approach.** *Biometrics* 2011, **67**:50–58. [http://dx.doi.org/10.1111/j.1541-0420.2010.01423.x]

27. Crowther MJ: **STJM: Stata module to fit shared parameter joint models of longitudinal and survival data.** *Stat Softw Components Boston Coll Dep Econ* 2012. [http://ideas.repec.org/c/boc/bocode/s457339.html]