RESEARCH

Open Access

Relationship between prostate-specific antigen, alkaline phosphatase levels, and time-to-tumor shrinkage: understanding the progression of prostate cancer in a longitudinal study

Madiha Liaqat[1], Rehan Ahmad Khan[1], Florian Fischer[2*] and Shahid Kamal[1]

Abstract

Background This study delves into the complex interplay among prostate-specific antigen, alkaline phosphatase, and the temporal dynamics of tumor shrinkage in prostate cancer. By investigating the longitudinal trajectories and time-to-prostate cancer tumor shrinkage, we aim to untangle the intricate patterns of these biomarkers. This understanding is pivotal for gaining profound insights into the multifaceted aspects of prostate cancer progression. The joint model approach serves as a comprehensive framework, facilitating the elucidation of intricate interactions among these pivotal elements within the context of prostate cancer .

Methods A new joint model under a shared parameters strategy is proposed for mixed bivariate longitudinal biomarkers and event time data, for obtaining accurate estimates in the presence of missing covariate data. The primary innovation of our model resides in its effective management of covariates with missing observations. Built upon established frameworks, our joint model extends its capabilities by integrating mixed longitudinal responses and accounting for missingness in covariates, thus confronting this particular challenge. We posit that these enhancements bolster the model's utility and dependability in real-world contexts characterized by prevalent missing data. The main objective of this research is to provide a model-based approach to get full information from prostate cancer data collected with patients' baseline characteristics (*Age*, body mass index (*BMI*), *GleasonScore*, *Grade*, and *Drug*) and two longitudinal endogenous covariates (*Platelets* and *Bilirubin*).

Results The results reveal a clear association between prostate-specific antigen and alkaline phosphatase biomarkers in the context of time-to-prostate cancer tumor shrinkage. This underscores the interconnected dynamics of these key indicators in gauging disease progression.

Conclusions The analysis of the prostate cancer dataset, incorporating a joint evaluation of mixed longitudinal prostate-specific antigen and alkaline phosphatase biomarkers alongside tumor status, has provided valuable insights into disease progression. The results demonstrate the effectiveness of the proposed joint model, as evidenced by accurate estimates. The shared variables associated with both longitudinal biomarkers and event

*Correspondence:
Florian Fischer
florian.fischer1@charite.de
Full list of author information is available at the end of the article

Liaqat *et al. BMC Urology*   (2024) 24:137

Page 2 of 8

times consistently deviate from zero, highlighting the robustness and reliability of the model in capturing the complex dynamics of prostate cancer progression. This approach holds promise for enhancing our understanding and predictive capabilities in the clinical assessment of prostate cancer.

Keywords Prostate cancer, Biomarker, Prostate-specific antigen, Alkaline phosphatase, Joint model

Background

Longitudinal data from biomarkers are usually collected in cancer studies, with the primary outcome variable that consists of time until the pre-specified event occurs, e.g., time to disease progression, recurrence of cancer, or death. Joint models are common to simultaneously analyze longitudinal and event time outcomes by addressing complex issues of informative dropouts, informative censoring, missingness, and association structure [1, 2]. One perspective is to directly introduce longitudinal measurements into Cox's model as time-varying covariates. However, the risk of bias increases in this approach if observations are measured with error or if there exists intra-individual variability [3].

Two further types of joint modeling approaches have been employed in previous studies: One approach utilizes the factorization of two outcomes as the product of marginal and conditional distributions, and the second approach utilizes random effects to account for the association between outcomes [4, 5]. When multiple outcomes are collected on individuals, the joint model is extended to allow for the correlation structure between different outcomes. More than one similar longitudinal with more than one event time outcome has been analyzed using the usual joint modeling strategy [6].

Simultaneous analysis of multivariate longitudinal and event time outcomes becomes complex to execute when different types of longitudinal outcomes are included in the study [7, 8]. Random-effects modeling strategy is the most frequently applied joint modeling approach, which is well established to analyze continuous and discrete longitudinal data implying an association structure between repeated measures on individuals [9].

Linear mixed-effects models commonly assume longitudinal processes [10, 11] and the proportional hazards model is specified for an event time process [12]. Joint models of multiple outcome processes are connected with parameters using various associated structures such as time-dependent slopes, interaction and lagged effects, cumulative effects, and random-effects parameterization [13].

The presence of missing data leads to a high complexity in longitudinal data due to different follow-up time points. Deletion of missingness has threats related to loss of information and reduction in precise estimates, which produces misleading inferences. The validity of accommodating incompleteness during data analysis depends upon the missingness mechanism. According to Rubin [14], it is based on missingness-completely-at-random (MCAR), missingness-at-random (MAR), and missingness-not-at-random (MNAR) taxonomy, by making assumptions related to observed and unobserved information. MCAR probability does not depend on unobserved or observed information, MAR probability depends upon observed information but it is independent of unobserved information, while the occurrence of MNAR purely depends upon observed and unobserved information.

A single missing value in the data set is replaced with a calculated mean or median or the last observed value is used to fill space of unobserved value. This technique produces biased estimates in the case of MAR [15]. Handling missing data in statistical methodology with a joint evaluation of longitudinal and event time responses provides substantial effects regarding inferences. Selection, pattern-mixture, and shared-parameter models (SPM) are three well-established joint modeling frameworks that are used to apply different factorizations for the joint distribution of outcomes and missing data processes [16]. The SPM approach mostly applies with conditionally independent assumption between outcomes and missingness based on random- effects [17, 18].

Our work aims to develop a joint model for continuous and binary longitudinal responses and an event time outcome with missing covariates. Shared random effects are used to build a joint model for mixed longitudinal outcomes with an-event time outcome; by assuming the existence of sharing effects between longitudinal mixed models and an-event time model. Longitudinal and event time outcomes are conditionally independent of shared parameters and covariates. The ignorability condition fulfills if missingness in time-dependent covariates relates to observed observations but not to unobserved data. A distributional assumption for incomplete observed covariates is inevitable for likelihood inference, while no such assumption is needed for covariates without missingness [19]. We propose a joint model for prostate-specific antigen (*PSA*) and alkaline phosphatase (*ALP*) biomarkers, and time-to-tumor shrinkage status data collected on prostate cancer patients. The proposed model is based on patients' baseline characteristics as covariates with time-dependent endogenous covariates.

Liaqat *et al. BMC Urology*   (2024) 24:137

Page 3 of 8

Methods

Notation and definitions

The methodology is based on formulation of a joint model for bivariate longitudinal responses, one continuous and normally distributed and the other one binary. Let y_{1ij} and y_{2ij} be the continuous and binary components of jth outcome at time point t_{ij} where $i = 1,2,3,\ldots,n$. Let $y_i = (y_{1i}^\top, y_{2i}^\top)^\top$ be the bivariate longitudinal outcome vector for i, such that $y_{hi} = (y_{hi1}, y_{hi2}, y_{hi3}, \ldots, y_{hij})^\top$ where $h = 1,2, j = 1,2,\ldots,n_{hi}$ is an n_{hi} column vector for h^{th} longitudinal responses for subject i.

For bivariate responses, the generalized linear mixed effects model takes the form as,

$$E(y_{hi}|u_{hi}) = g_h(X_{hi}\beta_h + Z_{hi}u_{hi}), h = 1,2 \qquad (1)$$

Where, $g_h(.)$ denotes bivariate link functions for mixed data types. X_{hi} and β_h represent the $n_{hi} \times p_h$ design matrix of covariate values of fixed effects p_h effects. In the same way, Z_{hi} and u_{hi} denote a $n_{hi} \times q_1$ design matrix of covariates of q_h dimensional vector of normally distributed random effects with mean vector 0 and covariance-matrix Σ. It is assumed that elements of y_{hi} are independent conditional on u_{hi}.

The generalized linear mixed effects model (1) is written using an identity link for continuous response and a logit link for binary response [20], respectively as,

$$E(y_{1i}|u_{1i}) = X_{1i}\beta_1 + Z_{1i}u_{1i}, \qquad (2)$$

$$logit(\Pr(y_{2i} = 1)|u_{2i}) = X_{2i}\beta_2 + Z_{2i}u_{2i}. \qquad (3)$$

It is assumed that u_{1i} follows a normal distribution with zero mean vector and variance-covariance matrix Σ, and u_{2i} is proportional to u_{1i}. It is written

censoring criteria [21]. It is assumed that event time for i^{th} individual is associated with continuous and binary longitudinal responses via shared random effects u_{1i} and u_{2i}. Specifically, the hazard function for *ith* individual at time T_i is given by a proportional model, given as,

$$h_i(t|x_{3i}, U_{3i}) = h_0(t)\exp(x_{3i}^\top\beta_3 + U_{3i}), \qquad (4)$$

where $h_0(T_i)$ is the baseline hazard function, assumed to be parametric or left unspecified. X_{3i} is a p_3 dimensional vector of covariates with regression coefficients β_3.

Here U_{3i}, shared parameters are used to associate the random effects of longitudinal outcomes with the random effect of event time outcome.

The shared random effects joint model becomes,

$$U_{3i} = \Psi_1^\top u_{1i} + \Psi_2^\top u_{2i} + u_{3i}, \qquad (5)$$

where, $\Psi_1^\top, and \Psi_2^\top$ are associated parameters and u_{3i} is a normally distributed frailty term. Longitudinal outcome vector y_i and an-event time outcome T are conditionally independent on covariates and random-effects vector u. Let's combine observed data as,

$$\{y_{hi}(t_{ij}), X(t_{ij}), Z(t_{ij}), T_i^*, \Delta_i, i = 1,2,3,\ldots,n, j = 1,2,3,\ldots,n_{hi}, h = 1,2\}. \qquad (6)$$

The joint distribution of observed data based on conditional independence assumption is given for individuals as,

$$f(y_i, T_i^*, \Delta_i|u_i; \theta) = f(y_{1i}|u_{1i}; \beta_1, \sigma_{\epsilon_1}^2) \times f((y_{2i}|u_{2i}; \beta_2) \\ \times f(T_i^*, \Delta_i|u_i; \beta_3, \Psi, \sigma_3^2), \qquad (7)$$

where θ denotes the parameter vector and $f(\cdot)$ denotes the probability density function. The log-likelihood for the observed data is expressed by,

$$\sum_{i=1}^{N} log \int_{u_i} \prod_{i=1}^{n_{1i}} [\frac{1}{\sqrt{2\pi\sigma_\epsilon^2}}\exp\{-\frac{1}{2\sigma_{\epsilon_1}^2}(y_{1ij} - x_{1ij}\beta_1 - z_{1ij}u_{1i})^2\}] \times$$

$$\prod_{j=1}^{n_{2i}} \left[\left\{\frac{1}{1 + \exp(-x_{2ij}\beta_2 - z_{2ij}u_{2i})}\right\}^{y_{1ij}} \left\{\frac{\exp(-x_{2ij}\beta_2 - z_{2ij}u_{2i})}{1 + \exp(-x_{2ij}\beta_2 - z_{2ij}u_{2i})}\right\}^{1-y_{2ij}}\right]$$

$$\times [h_0(T_i^*)\exp\{x_{3i}^T\beta_3 + U_i\}]^{\Delta_i}\exp\{-\int_0^{T_i^*} h_0(s)\exp\{x_{3i}^T\beta_3 + U_i\}ds\} \times f(u_i|\Sigma)du_i. \qquad (8)$$

as $u_{2i} = A_0u_{1i}$, where A_0 denotes a diagonal matrix of unknown constants.

True event time T_i and censoring time C_i are defined for individual i, using observed event time $T_i^* = \min(T_i, C_i)$. Censoring status is represented with $\Delta_i = I(T_i \leq C_i)$, by assuming non-informative

If associated parameters $\Psi = (\Psi_1^\top, \Psi_2^\top)^\top = (0,0)$ are zero, there exists no association among the three outcome processes, in this case, joint modeling is not necessary and separate models can be used. In Bayesian inference, Markov-Chain-Monte-Carlo methods including Gibbs sampler and Metropolis-Hastings algorithm

Liaqat *et al. BMC Urology*   (2024) 24:137

Page 4 of 8

implemented to draw samples iteratively from conditional posterior distribution [22]. The conditional posterior distribution is written with joint prior distribution $p\left(\theta'_{y_{1i}}, \theta'_{y_{2i}}, \theta'_{T_i^*}, \theta'_{\Delta_i}, \theta'_{u_i}\right)$ of all parameters.

Let x_i^{mis} be a vector of time-dependent missing covariates included in the generalized linear model, which is given as,

$$p(x_{ij}|\theta_{ij}, \tau) = \exp\left\{\alpha_i^{-1}(\tau)(x\theta_{ij} - b(\theta_{ij})) + c(x, \tau)\right\}, x_{ij} \in x_i^{mis};$$ (9)

the term $\prod_{i \epsilon x_i^{mis}} \prod_{i=1}^n p(x_{ij}|\theta_{ij}, \tau)$ be added to the conditional posterior distribution.

Let x_{ij}^{mis} be missing covariates and x_{ij}^{obs} be observed covariates, and $\left\{\left(R_{ij}, y_{ij}, x_{ij}^{obs}, x_{ij}^{mis}\right)\right\}$ be full data, where R_{ij} is missing data indicator, which takes value 0 or 1, based on observe or missing observations, and θ is parameter of missingness. R_{ij} is independent of missing observations x_{ij}^{mis} but may depend on observed values (y_{ij}, x_{ij}^{obs}). We assume random-effects association for two heterogeneous longitudinal processes, this parameterization is more meaningful for our case, where random-intercept and random-slope are assumed for longitudinal sub-models. Both frequentist and Bayesian statisticians have agreed on the fact of inaccurate estimates in case of not incorporating missing covariates in data analysis [19]. In a frequentist context, many parameters of the likelihood function are not identifiable due to missing covariates, Bayesian analysis provides a solution to this problem by appropriately choosing proper prior distributions, using informative priors. Informative priors add information about unknown parameters when based on true available historical data. If no previous knowledge is available, non-informative prior distribution and Gibbs sampling are used to combine with the Metropolis-Hastings algorithm under full conditional distribution of each model parameter [23]. We implement Bayesian thinking for parameter inferences, using the Gibbs sampling algorithm with the help of 'OpenBUGS' [24]. Weak-informative priors are chosen, and fixed effects parameters are assumed to be distributed normally with large variance components. For variances of error terms inverse-gamma prior distributions and for covariances of random-effects inverse-Wishart distributions are applied as priors. The implementation code is available upon request from the corresponding author.

Application to mixed longitudinal (PSA, ALP) and a time-to-tumor status prostate cancer study

Prostate cancer being the major cause of cancer deaths in men has increased the disease burden in terms of diagnosis and progression to assess the effectiveness of treatment. In the era from 1988 to the mid-1990s, *PSA* was the only serum marker for prostate cancer and it is still used for diagnosis, measuring treatment efficacy, and recurrence of prostate cancer [25, 26]. *ALP* is one of the older biomarkers for investigating and monitoring prostate cancer. Furthermore, it is a reliable indicator of bone metastases [27, 28]. Various authors have proved an increase in the clinical effectiveness of simultaneous measurement of *PSA* and *ALP* in prostate cancer patients [29, 30].

Our motivation for joint modeling of longitudinal and event time outcomes arises from a prostate cancer dataset collected from one of the most renowned public hospitals in Pakistan. This dataset includes patients who underwent various treatments and were followed to monitor PSA and ALP measurements until tumor shrinkage or censoring occurred. To study the effects of biomarkers on event time, we jointly modeled longitudinal PSA and ALP biomarkers along with tumor shrinkage outcomes.

The study included patients diagnosed with prostate cancer as their primary disease, who provided blood samples during follow-up visits spaced 28 to 30 days apart. The demographic characteristics of the study sample showed an age distribution ranging from 50 to 85 years, with a mean age of 68 years. Clinically, the patients varied in disease stage, with a significant number in advanced stages (stages III and IV), indicating a high-risk population.

Treatment histories of the patients included a mix of therapeutic approaches: surgery (prostatectomy), radiation therapy, hormone therapy (androgen deprivation therapy), and combinations thereof. This diversity in treatment history reflects the standard care practices for managing advanced prostate cancer and provides a comprehensive view of the patient population under investigation.

To remove variations caused by different scales of data, we standardized all continuous variables. Specifically, each variable was transformed by subtracting its mean and dividing by its standard deviation, resulting in variables with a mean of zero and a standard deviation of one. This standardization ensures that all continuous variables contribute equally to the model and prevents those with larger scales from dominating the analysis.

Additional preprocessing steps included addressing missing data through multiple imputations to mitigate bias from incomplete datasets. Outliers were identified and assessed, with extreme values either transformed or excluded based on their influence on the overall analysis. Categorical variables were encoded using one-hot encoding to convert them into a format suitable for modeling.

Results

To jointly model three outcomes the model is specified as,

$$y_{1ij} = \eta_{1i}(t_{ij}) + U_{1i}(t_{ij}) + \epsilon_{ij1}. \tag{10}$$

To model longitudinal $log(PSA)$ a linear-mixed-effects model with random intercept and slope is specified as,

$$\log(PSA_{ij}) = \beta_{11} + \beta_{12}t_{ij} + \beta_{13}Age_i + \beta_{14}Platelets_i + \beta_{15}BMI_i +$$

$$\beta_{16}Bilirubin_i + \beta_{17}GleasonScore_i + \beta_{18}Grade_i + \beta_{19}Drug_i$$

$$+ u_{10i} + u_{11i}t_{ij} + \epsilon_1(t). \tag{11}$$

The random-intercept u_{10} and random-slope u_{11} are assumed to follow a bivariate normal distribution with means 0 and variance-covariance components $\sigma^2_{u\,10}, \sigma^2_{u\,11}$, and $\sigma_{u_{10}u_{11}}$. In addition, the random error term $\epsilon_{1ij} \sim N(0, \sigma^2_{\epsilon1})$ is independent of random effects.

Binary longitudinal outcome (*ALP*) considers a logistic mixed-effects regression model as,

$$logit(P(y_{2ij} = 1)) = \eta_{2i}(t_{ij}) + U_{2i}(t_{ij}), \tag{12}$$

where,

$$= \beta_{21} + \beta_{22}s_{ij} + \beta_{23}Age_i + \beta_{24}Platelets_i + \beta_{25}BMI_i +$$
$$\beta_{26}Bilirubin_i + \beta_{27}GleasonScore_i + \beta_{28}Grade_i$$
$$+ \beta_{29}Drug_i + \Psi_1u_{10i} + \Psi_2u_{11i}t_{ij}. \tag{13}$$

The third outcome is time to tumor shrinkage. A Weibull model is proposed such that random- effects are shared among three sub-models,

$$T_i \sim \eta_{3i} + U_{3i}, \tag{14}$$

$$\eta_{3i} = \exp(\beta_{31} + \beta_{32}GleasonScore_i + \beta_{33}Grade_i$$
$$+ \beta_{34}Drug_i + \Psi_3u_{10i} + \Psi_4u_{11i} + u_{3i}). \tag{15}$$

Considering three sub-models under SPM, random-effects terms are defined as follows:

$$U_{1i}(t_{ij}) = u_{10i} + u_{11i}t_{ij}, \tag{16}$$

$$U_{2i}(t_{ij}) = \Psi_1u_{10i} + \Psi_2u_{11i}t_{ij}, \tag{17}$$

$$U_{3i} = \Psi_3u_{10i} + \Psi_4u_{11i} + u_{3i}, \tag{18}$$

It is to be assumed that $\epsilon_{ij} \sim N(0, \sigma^2_\epsilon)$, $\epsilon_{ij\prime s}$ indicates error components. The regression coefficients $\beta_{11}, \ldots, \beta_{19}, \beta_{21}, \ldots, \beta_{29}$, and $\beta_{31}, \ldots, \beta_{34}$ are fixed effects unknown parameters. In these sub-models, Age_i is the baseline age of patients in years, $Platelets_i$ counts

are recorded over time, BMI_i of patients are recorded at entry-level, and $Bilirubin_i$ measurements are taken at each time of measurement. $GleasonScore_i$, $Grade_i$, and $Drug_i$ are binary covariates.

Estimation and inference of our proposed joint model are based on conditional independence assumptions. For our analysis, the two longitudinal biomarkers constitute the joint model by adding additional variance-covariance parameters of the random effects. It is also assumed

Table 1 Parameter estimates (Est.), standard deviation (S.D.) and 95% credible interval (CI) of PCa data, by applying our proposed joint model

Parameter	Est.	S.D.	95% CI
log(PSA) biomarker			
Intercept	0.900	0.100	0.800, 1.100
Obstime	-7.000	0.100	-7.200, -6.800
Age	0.200	0.000	0.200, 0.300
Platelets	0.200	0.000	0.200, 0.300
Body mass index (BMI)	-0.100	0.000	-0.200, -0.100
Bilirubin	0.100	0.000	0.000, 0.200
Gleason Score	1.900	1.300	0.400, 3.500
Grade	-0.900	1.300	-2.500, 0.800
Drug	0.400	0.200	0.100, 0.700
σ^2_ϵ	1.100	0.000	1.100, 1.200
$\sigma^2_{u_1}$	0.400	0.000	0.400, 0.500
***ALP* biomarker**			
Intercept	1.900	0.300	1.300, 2.600
Obstime	1.500	0.800	0.200, 3.200
Age	-0.300	0.100	-0.500, -0.100
Platelets	0.500	0.100	0.300, 0.800
Body mass index (BMI)	-0.300	0.100	-0.500, -0.200
Bilirubin	4.800	0.500	3.900, 5.700
Gleason Score	4.000	1.600	1.600, 7.200
Grade	-2.500	1.700	-5.700, 0.600
Drug	-1.100	0.600	-2.500, 0.200
σ^2_{u2}	0.600	0.500	0.100, 1.700
$\sigma_{u_1u_2}$	0.000	0.100	-0.200, 0.100
Time-to-tumor shrinkage			
Intercept	-19.400	0.400	-20.300, -18.600
Gleason Score	0.200	0.100	0.100, 0.300
Grade	-0.400	0.100	-0.600, -0.100
Drug	0.200	0.100	0.100. 0.400
σ^2_u	0.300	0.200	0.100, 0.700
r	3.500	0.100	3.300, 3.600
tau	0.900	0.000	0.800, 0.900
Ψ_1	-0.100	0.200	-0.400, 0.200
Ψ_2	-10.100	5.600	-23.600, -3.100
Ψ_3	-0.500	0.400	-1.100, 0.100
Ψ_4	-0.600	0.300	-1.400, -0.100

Liaqat *et al. BMC Urology*   (2024) 24:137

Page 6 of 8

that given the observed history, the right censoring and observation time of visits are independent of the true event times and future biomarkers' measurements.

The parameter estimates (Est.), standard deviation (S.D.), and 95% credible intervals (CI) for the prostate cancer dataset were obtained using the proposed joint model, focusing on longitudinal PSA and ALP biomarkers and time-to-tumor shrinkage outcomes. These results, summarized in Table 1, provide insights into the magnitude and direction of associations between covariates and prostate cancer progression.

Combined analysis of this study aimed to identify the association structure among the three models (linear mixed-effects for log(*PSA*), logistic mixed-effects for *ALP*, and event of interest "shrinkage of tumor"), which is directly assessed from the associated parameters mentioned in Table 1.

The positive intercepts for both *PSA* (0.900) and *ALP* (1.900) biomarkers suggest baseline levels before considering other covariates. *PSA* levels decrease significantly over time (Est. = -7.000), while *ALP* levels show an increase (Est. = 1.500), reflecting different temporal dynamics. Higher platelet counts are associated with increases in both *PSA* (Est. = 0.200) and *ALP* (Est. = 0.500). Elevated bilirubin shows a slight increase in *PSA* (Est. = 0.100) but a substantial increase in *ALP* (Est. = 4.800). Higher *GleasonScores* significantly increase both *PSA* (Est. = 1.900) and *ALP* (Est. = 4.000) levels. *Drug* treatment is associated with increases in *PSA* (Est. = 0.400) and decreases in *ALP* (Est. = -1.100).

The first associated parameter estimated value is -0.100 which suggests a slight negative effect, although not statistically significant, with the 95% CI ranging from -0.400 to 0.200, indicating uncertainty about the true effect size. The second associated parameter indicates a more substantial effect compared to the first associated parameter with an estimated value of -10.100, which suggests a significant negative impact, with the 95% CI ranging from -23.600 to -3.100, indicating a wide range of possible effect sizes. The estimated value of the third associated parameter Ψ_3 is -0.500 which suggests a negative effect, but with a wide 95% CI (-1.100, 0.100), which indicates uncertainty about the true effect size and direction. The estimated value of Ψ_4 is -0.600 which suggests a moderate negative impact, with a relatively narrow 95% CI from -1.400 to -0.100, providing more confidence in the estimated effect size.

Results indicate that *PSA* is a good biomarker in the diagnosis of prostate cancer, but after treatment, it shows different trends for different treatments. *PSA* values point out successful or unsuccessful prostate cancer treatment. An increased level of *ALP* in prostate cancer may be an indicator of advanced cancer, as it means cancer is spreading to bones and tissues. That situation is alarming and emphasizes physicians to make decisions regarding advanced therapies and sometimes combinations of androgen deprivation therapy with prostatectomy. The association between *PSA* and *ALP* emphasizes taking into account these two biomarkers simultaneously to get insights into prostate cancer progression.

Discussion

We developed a joint model for *PSA* and *ALP* biomarkers and a time-to-tumor shrinkage variable and applied this model to analyze the prostate cancer dataset to find any potential association between biomarkers and tumor status accounting for patients' baseline characteristics and endogenous *Platelets* and *Bilirubin* covariates. We applied a joint modeling approach using parametric models for the longitudinal biomarkers, and Cox's proportional hazards model for time-to-tumor shrinkage. We proposed a joint model by employing a Bayesian framework, including posterior estimates, CI, and DIC model selection criteria. Results from the fitting model indicate that *PSA* and *ALP* biomarkers jointly contribute significantly to prostate cancer tumor shrinkage. Prior distributions are chosen taking inspiration from Choi et al. [23].

Results depict the existence of associations among the different outcomes, so separate models may provide inconsistent estimates. Efficiency gain by employing a joint model as compared to separate models for mixed types of longitudinal (continuous (*PSA*) and binary (*ALP*)) outcomes, which are strongly associated with each other, means these outcome processes are dependent. We focus on parametric Cox's model to analyze right-censored event time data by applying Weibull distribution for Cox's model. Joint numerical analysis of heterogeneous data is very difficult and computationally challenging, complexity arises due to the existence of missingness in time-dependent covariates (*Platelets* and *Bilirubin*). Researchers usually apply list-wise deletion or imputation techniques to handle missing data, which has shortcomings of not providing complete information. In this regard, we provide a novel contribution by formulating a joint model for mixed types of responses, incorporating missingness in time-dependent covariates. More importantly, we provide insights into the role of association parameters in joint modeling, very little literature is available on this line of investigation.

It is recommended that prostate cancer patients should be assessed for *PSA* observations during the first three months after treatment, in some cases *PSA* measurements become undetectable by the first month after prostatectomy. *PSA* alone is not an adequate biomarker for prostate cancer progression [31], *ALP* is another good biomarker that is correlated with *PSA*. Correlation of both biomarkers

Liaqat *et al. BMC Urology* (2024) 24:137

Page 7 of 8

confirmed that prostate cancer patients should observed for *PSA* and *ALP* simultaneously after treatment to confirm treatment effectiveness on tumor shrinkage.

Finally, we utilized shared random effects to characterize the relationship between longitudinal biomarkers. A potential shortcoming of the shared random effects joint modeling approach is that the structure of random effects for longitudinal biomarkers is limited. Other joint model formulations are available as an alternative to shared random effects like pattern mixture and, selection modeling approaches, those are employed and results can be compared.

Joint modeling of longitudinal and event time has previously been applied to get insights into different diseases, the novelty of this research is that we employed a joint model for two prostate cancer longitudinal biomarkers (*PSA* and *ALP*), and time to tumor shrinkage. This research aims to provide insight into prostate cancer progression by monitoring *PSA* and *ALP* for multivariate covariates until tumor status reaches a satisfactory level according to physicians' directives. The joint modeling approach is an improvement over traditional separate longitudinal and event time models. This study illustrates the usefulness of simultaneous modeling for analyzing the prostate cancer dataset, but the approach is also applicable to a wide variety of diseases, where multiple mixed types of longitudinal responses may have an association with event time response.

SPMs assume that random effects follow distributions such as normal, gamma, or t-distribution to capture individual-specific characteristics in longitudinal outcomes. Departures from these assumptions can introduce bias into estimates. Error terms in these models often assume Gaussian or skewed distributions like gamma or log-normal [32]. Accurate modeling of within-individual correlations, such as autoregressive or compound symmetry structures, is essential to avoid bias. Longitudinal outcome relationships with time can be modeled using parametric or non-parametric approaches, affecting trend interpretation and prediction accuracy. While these assumptions simplify modeling and estimation, violations can lead to bias, necessitating sensitivity analyses for robustness. Limitations of these models include requirements for large sample sizes, computational challenges, and potential difficulties in interpretation, with sensitivity analyses playing a crucial role in identifying sources of bias and enhancing understanding.

In the future, this joint model can be extended to incorporate additional biomarkers and clinical variables to provide a more comprehensive understanding of prostate cancer progression and treatment response. In addition, the inclusion of genetic markers or imaging data could enhance the predictive accuracy of the model.

Advanced machine learning techniques can also be employed within the joint modeling framework. Techniques such as deep learning could capture complex, non-linear relationships between biomarkers and clinical outcomes can potentially lead to more accurate and personalized predictions. Additionally, our future research plan is to investigate the impact of different imputation methods for missing data, including multiple imputation and machine learning-based approaches to enhance the robustness of the findings.

Abbreviations

ALP	Alkaline phosphatase
BMI	Body mass index
CI	Credible interval
MAR	Missingness-at-random
MCAR	Missingness-completely-at-random
MNAR	missingness-not-at-random
PSA	Prostate-specific antigen
S.D.	Standard deviation
SPM	Shared-parameter models

Authors' contributions
ML conceived the original idea of the study, designed the study, analyzed the data and drafted the manuscript. SK supervised the whole study design. RAK, FF and SK revised the manuscript critically for important intellectual content. All authors approved the final version of the manuscript.

Availability of data and materials
Data is available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
The study is based on a retrospective analysis of secondary analysis of administrative data. Mayo Hospital informed patients that their collected data might be used for research purposes. Therefore, no informed consent to participate is needed as approved by the ethical committee of University of the Punjab. Personal information was kept confidential to others during data collection from reports. The study has been conducted in accordance with the Declaration of Helsinki. The study was approved by the College of Statistical and Actuarial Sciences Research Board and the ethical committee at the University of the Punjab, Lahore Pakistan.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]College of Statistical and Actuarial Sciences (CSAS), University of the Punjab, Lahore, Pakistan. [2]Institute of Public Health, Charité – Universitätsmedizin Berlin, Berlin, Germany.

Liaqat *et al. BMC Urology*   (2024) 24:137

Page 8 of 8

References

1. Li N, Elashoff RM, Li G, Saver J. Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial. Stat Med. 2010;29(5):546–57.
2. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. Biometrics. 2011;67(3):819–29.
3. Sattar A, Sinha SK. Joint modeling of longitudinal and survival data with a covariate subject to a limit of detection. Stat Methods Med Res. 2019;28(2):486–502.
4. Catalano PJ, Ryan LM. Bivariate latent variable models for clustered discrete and continuous outcomes. J Am Stat Assoc. 1992;87(419):651–8.
5. Fieuws S, Verbeke G. Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. Stat Med. 2004;23(20):3093–104.
6. Li K, Luo S. Bayesian functional joint models for multivariate longitudinal and time-to-event data. Comput Stat Data Anal. 2019;129:14–29.
7. Liu X, Daniels MJ, Marcus B. Joint models for the association of longitudinal binary and continuous processes with application to a smoking cessation trial. J Am Stat Assoc. 2009;104(486):429–38.
8. Ivanova A, Molenberghs G, Verbeke G. Mixed models approaches for joint modeling of different types of responses. J Biopharm Stat. 2016;26(4):601–18.
9. Gao S. A shared random effect parameter approach for longitudinal dementia data with non-ignorable missing data. Stat Med. 2004;23(2):211–9.
10. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. Biometrics. 1997;53(1):330–9.
11. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. Biostatistics. 2000;1(4):465–80.
12. Cox DR. Regression models and life-tables. J Royal Stat Soc Ser B (Methodological). 1972;34(2):187–220.
13. Rizopoulos D, Hatfield LA, Carlin BP, Takkenberg JJ. Combining dynamic predictions from joint models for longitudinal and time-to-event data using bayesian model averaging. J Am Stat Assoc. 2014;109(508):1385–97.
14. Rubin DB. Inference and missing data. Biometrika. 1976;63(3):581–92.
15. Woolley SB, Cardoni AA, Goethe JW. Last-observation-carried-forward imputation method in clinical efficacy trials: review of 352 antidepressant studies. Pharmacotherapy. 2009;29(12):1408–16.
16. Wu L, Hu XJ, Wu H. Joint inference for nonlinear mixed-effects models and time to event at the presence of missing data. Biostatistics. 2008;9(2):308–20.
17. Vonesh EF, Greene T, Schluchter MD. Shared parameter models for the joint analysis of longitudinal data and event times. Stat Med. 2006;25(1):143–63.
18. Papageorgiou G, Rizopoulos D. An alternative characterization of MAR in shared parameter models for incomplete longitudinal data and its utilization for sensitivity analysis. Stat Modelling. 2021;21(1–2):95–114.
19. Ibrahim JG, Chen MH, Lipsitz SR. Bayesian methods for generalized linear models with covariates missing at random. Can J Stat. 2002;30(1):55–78.
20. Parzen M, Ghosh S, Lipsitz S, Sinha D, Fitzmaurice GM, Mallick BK, Ibrahim JG. A generalized linear mixed model for longitudinal binary data with a marginal logit link function. Ann App Stat. 2011;5(1):449.
21. Alam K, Maity A, Sinha SK, Rizopoulos D, Sattar A. Joint modeling of longitudinal continuous, longitudinal ordinal, and time-to-event outcomes. Lifetime Data Anal. 2021;27:64–90.
22. Chi YY, Ibrahim JG. Joint models for multivariate longitudinal and multivariate survival data. Biometrics. 2006;62(2):432–45.
23. Choi J, Anderson SJ, Richards TJ, Thompson WK. Prediction of transplant-free survival in idiopathic pulmonary fibrosis patients using joint models for event times and mixed multivariate longitudinal data. J Appl Stat. 2014;41(10):2192–205.
24. Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: evolution, critique and future directions. Stat Med. 2009;28(25):3049–67.
25. Oesterling JE. Prostate specific antigen: a critical assessment of the most useful tumor marker for adenocarcinoma of the prostate. J Urol. 1991;145(5):907–23.
26. Moul JW. Prostate specific antigen only progression of prostate cancer. J Urol. 2000;163(6):1632–42.
27. Bishop MC, Hardy JG, Taylor MC, Wastie ML, Lemberger RJ. Bone imaging and serum phosphatases in prostatic carcinoma. Br J Urol. 1985;57(3):317–24.
28. Rao SR, Snaith AE, Marino D, Cheng X, Lwin ST, Orriss IR, Hamdy FC, Edwards CM. Tumour-derived alkaline phosphatase regulates tumour growth, epithelial plasticity and disease-free survival in metastatic prostate cancer. Br J Cancer. 2017;116(2):227–36.
29. Wang ZL, Wang XF. [Relationship of serum prostate-specific antigen and alkaline phosphatase levels with bone metastases in patients with prostate cancer]. Natl J Androl. 2005;11(11):825–7.
30. Lorente JA, Morote J, Raventos C, Encabo G, Valenzuela H. Clinical efficacy of bone alkaline phosphatase and prostate specific antigen in the diagnosis of bone metastasis in prostate cancer. J Urol. 1996;155(4):1348–51.
31. Liaqat M, Kamal S, Fischer F. Illustration of association between change in prostate-specific antigen (PSA) values and time to tumor status after treatment for prostate cancer patients: a joint modelling approach. BMC Urol. 2023;23:202.
32. El Saeiti R, García-Fiñana M, Hughes DM. The effect of random-effects misspecification on classification accuracy. Int J Biostat. 2021;18:279–92.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.