

Cite this article as: Wang X, Andrinopoulou E-R, Veen KM, Bogers AJC, Takkenberg JJM. Statistical primer: an introduction to the application of linear mixed-effects models in cardiothoracic surgery outcomes research—a case study using homograft pulmonary valve replacement data. Eur J Cardiothorac Surg 2022; doi:10.1093/ejcts/ezac429.

# Statistical primer: an introduction to the application of linear mixed-effects models in cardiothoracic surgery outcomes research—a case study using homograft pulmonary valve replacement data

Xu Wang<sup>a</sup>, Eleni-Rosalina Andrinopoulou<sup>b,c</sup>, Kevin M. Veen<sup>a</sup>, Ad J.J.C. Bogers <sup>a</sup> and Johanna J.M. Takkenberg<sup>a,\*</sup>

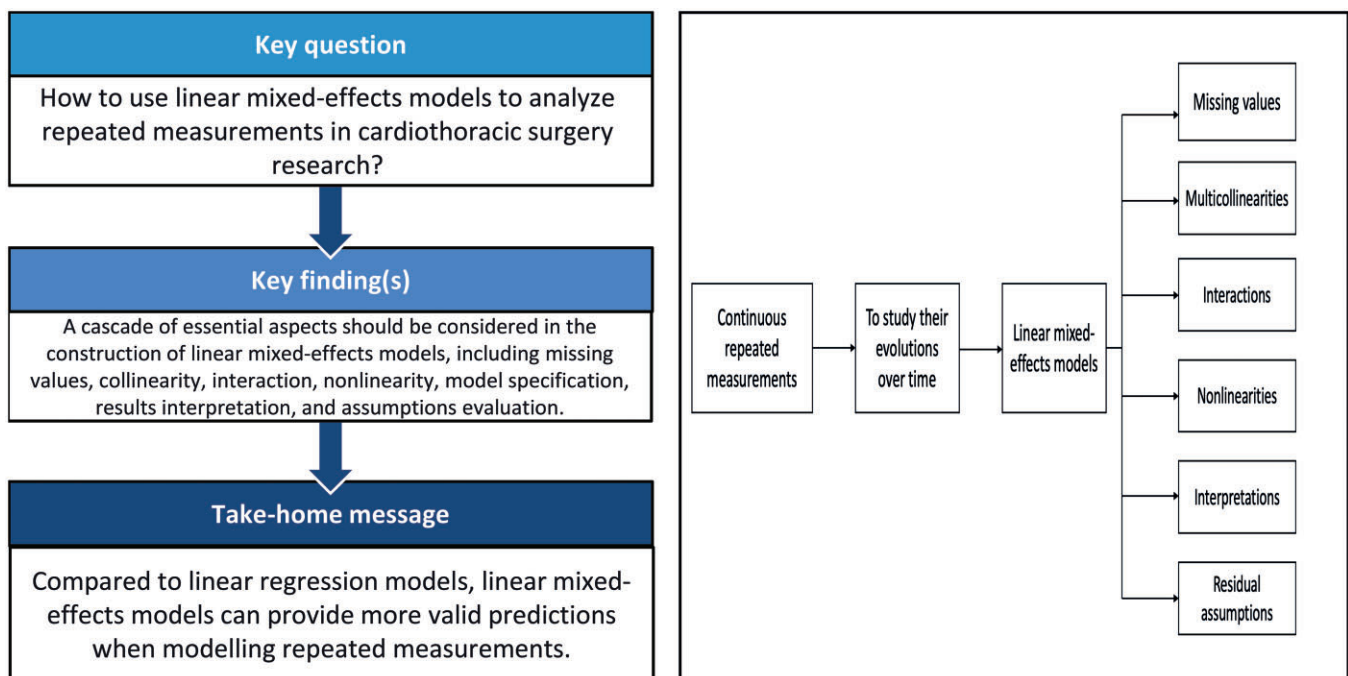
<sup>a</sup> Department of Cardiothoracic Surgery, Erasmus University Medical Center, University Medical Center Rotterdam, Rotterdam, Netherlands

<sup>b</sup> Department of Biostatistics, Erasmus University Medical Center, University Medical Center Rotterdam, Rotterdam, Netherlands

<sup>c</sup> Department of Epidemiology, Erasmus University Medical Center, University Medical Center Rotterdam, Rotterdam, Netherlands

\* Corresponding author. Department of Cardiothoracic Surgery, Erasmus University Medical Centre, Room Rg-633, PO Box 2040, 3000 CA Rotterdam, Netherlands. Tel: +31-10-7035413; e-mail: j.j.m.takkenberg@erasmusmc.nl (J.J.M. Takkenberg).

Received 4 January 2022; received in revised form 29 July 2022; accepted 23 August 2022



## Summary

**OBJECTIVES:** The emergence of big cardio-thoracic surgery datasets that include not only short-term and long-term discrete outcomes but also repeated measurements over time offers the opportunity to apply more advanced modelling of outcomes. This article presents a detailed introduction to developing and interpreting linear mixed-effects models for repeated measurements in the setting of cardiothoracic surgery outcomes research.

**METHODS:** A retrospective dataset containing serial echocardiographic measurements in patients undergoing surgical pulmonary valve replacement from 1986 to 2017 in Erasmus MC was used to illustrate the steps of developing a linear mixed-effects model for clinician researchers.

**RESULTS:** Essential aspects of constructing the model are illustrated with the dataset including theories of linear mixed-effects models, missing values, collinearity, interaction, nonlinearity, model specification, results interpretation and assumptions evaluation. A comparison between linear regression models and linear mixed-effects models is done to elaborate on the strengths of linear mixed-effects models. An R script is provided for the implementation of the linear mixed-effects model.

**CONCLUSIONS:** Linear mixed-effects models can provide evolutionary details of repeated measurements and give more valid estimates compared to linear regression models in the setting of cardio-thoracic surgery outcomes research.

**Keywords:** Mixed-effects model • Pulmonary valve replacement • Homograft

## ABBREVIATIONS

MNAR	Missing not at random
PVR	Pulmonary valve replacement
RVOT	Right ventricular outflow tract

## INTRODUCTION

Many patients with operated congenital heart disease (CHD) reach adulthood nowadays and have a good life expectancy [1, 2]. Pulmonary valve replacement (PVR) is one of the common procedures performed in patients with CHD. Because of the inevitable degeneration or dysfunction of currently available biological valve prostheses, patients undergoing PVR need to be followed up regularly to monitor valve function. Echocardiography is a commonly used tool in evaluating valve prostheses function [3]. Usually, multiple serial echocardiographic measurements will be performed in a patient before valve function starts to deteriorate. Different from independent measurements, repeated measurements in the same patient are dependent. A linear regression model or logistic regression model is not suitable in this case, due to the violation of the independence assumption. With inappropriate statistics, the clinical relevancy of the study might be undermined, even misdirecting the treatment decisions, which could harm patients' health. This is why extra caution should be given to the selection of optimal methodology to assess valve function over time. In the following part, one of the commonly applied models tackling repeated measurements is illustrated.

Let us consider a 21-year-old female patient who underwent surgical PVR with a pulmonary homograft for isolated severe pulmonary valve stenosis. Serial echocardiography was employed to monitor the function of the homograft. The patient had no prior cardiac surgery, and no concomitant procedures were done at the time of PVR. To date, she has been followed up for 7 years and the echocardiographic assessment of her right ventricular outflow tract (RVOT) peak gradient was measured repeatedly, in total 6 times. Her clinician and the patient are eager to know what will happen to her homograft 5–10 years from now, and when she perhaps would require a reintervention.

Traditionally, time-to-event or survival analyses can be used to predict time to reintervention based on baseline patient characteristics. However, survival analysis has obvious limitations in predicting valve function [4]. Valve function is a dynamic process, and it is necessary to account for all available measurements to draw the evolutionary curve, rather than one 'snapshot' of these measurements [4]. For this purpose, a regression model can be used. Which regression model to select depends on the characteristics of these measurements. Linear

regression models are suitable for continuous variables, for example, RVOT peak gradient in mmHg, and logistical regression models can be applied to categorical variables, e.g. pulmonary valve failure or not. Therefore, a linear regression model might be a solution. However, traditional linear regression models are not appropriate to analyse repeated measurements because the independence assumption is violated [5]. For example, the 6 echocardiographic measurements in our 21-year-old female are not independent and each measurement could influence the following one measured.

To address this, a mixed-effects model is more suitable as this model takes the individual echocardiographic trajectories into account, with more valid estimates of RVOT peak gradient over time. There are 2 main types of mixed-effects models: linear mixed-effects models for continuous variables and generalized mixed-effects models for categorical variables (binary, multinomial or ordinal response) [6]. These 2 types of models have similar procedures of construction. Only the linear mixed-effects model will be introduced in this article since the repeated measurements of the example dataset concern continuous outcome data. Although the name linear mixed-effects models suggest that these models only address linear associations, it should be stressed that they can also account for nonlinearity between response variables and covariates. That means that the applicability of linear mixed-effects models depends on the form of response variable. Nonlinearity can be explained in linear mixed-effects models by introducing, e.g. transformation, splines or polynomial. More details of the nonlinear relationship will be elaborated on in the 'Nonlinearity' part of this article.

A detailed introduction to developing and interpreting linear mixed-effects models in the setting of cardiothoracic surgery outcomes research is presented with the use of a dataset concerning patients whose RVOT peak gradients were followed echocardiographically after homograft PVR.

## MATERIALS AND METHODS AND RESULTS

### Ethics statement

This study was approved by the Medical Ethics Review Committee in Erasmus Medical Center (MEC 12-477). Individual informed consents were waived because of anonymity.

### Introduction of the dataset

All patients operated upon for elective homograft PVR at the Erasmus University Medical Center from April 1986 to November 2017 were identified retrospectively. In total, 701 homografts

**Table 1:** Baseline characteristics of patients with homograft implantation in right ventricular outflow tract

Parameters, median (IQR) or n (%)	Information (n = 623)
Age (years)	18.96 (8.05, 30.86)
Sex	
Female, n (%)	251 (40.29)
Male, n (%)	372 (59.71)
Height (cm)	163.00 (124.00, 175.00)
Weight (kg)	55.00 (23.00, 70.00)
Homograft type	
Aortic, n (%)	64 (10.56)
Pulmonary, n (%)	542 (89.44)
Diameter	23.00 (22.00, 25.00)
CPB time (min)	144.00 (96.00, 190.00)
Original diagnoses	
TOF	217 (34.83)
Aortic valve disease	175 (28.09)
PA with/without VSD	88 (14.13)
Others	143 (22.95)
Concomitant procedures	
With	459 (76.25)
Without	143 (23.75)
Proximal graft connection <sup>a</sup>	
With	102 (16.83)
Without	504 (83.17)
Distal graft connection <sup>a</sup>	
With	20 (3.29)
Without	587 (96.71)
No. previous heart operation	
0	117 (20.71)
1	246 (43.54)
2	137 (24.25)
≥3	65 (11.50)
No. implanted homograft	
First	530 (85.07)
Second	85 (13.64)
Third	6 (0.97)
Fourth	2 (0.32)

<sup>a</sup>Whether homograft has extra graft at its proximal or distal end to connect it with right ventricle or pulmonary artery.

CPB: cardiopulmonary bypass; IQR: interquartile range; PA: pulmonary atresia; TOF: tetralogy of Fallot; VSD: ventricular septal defect.

were implanted in 603 patients. Only patients who survived until discharge and had at least 1 postoperative echocardiographic examination containing RVOT peak gradient were included in the analyses. Thus, 623 homografts implanted in 537 patients met the inclusion criteria and were analysed. These patients had a total of 5111 measurements (median: 7; range: 1–26) and the mean follow-up was 11.46 years. The RVOT peak gradient was collected as a continuous variable (mmHg) and all measurements within a patient were treated as repeated measurements and analysed as outcomes. Baseline variables are presented in Table 1.

## Theory and application of linear mixed-effects models

Unlike linear regression models, linear mixed-effects models account for ‘intra-correlations’ between within-subject repeated measurements. Let us compare the example of the 21-year-old female above to a 45-year-old male who underwent PVR for

regurgitation after tetralogy of Fallot correction. It can be expected that the time-related pattern of echocardiographic measurements in the 45-year-old male is different from the 21-year-old female, as we know that there is an association between younger patient age and degeneration rate [7]. Linear mixed-effects models take these intra-correlations into account.

There are 2 parts in linear mixed-effects models: the random-effects part and the fixed-effects part. The random-effects part accounts for the intra-correlations of repeated measurements within each patient. The fixed-effects part accounts for the impact of various covariates on outcomes on an average level. These 2 parts play different roles in linear mixed-effects models and work together to give valid estimates.

In practice, it is very common to have sampling challenges, censoring (truncation) by death, unequal number of repeated measurements of the outcome per patient and variability in time among repeated measurements (such as serial echocardiographic assessment at different intervals after treatment). All of these could generate unbalanced longitudinal data. Linear mixed-effects models are very flexible in dealing with such unbalanced continuous longitudinal data. Moreover, it can provide predictions on not only a population level but also a patient-specific level, which is a big advantage compared to other approaches, e.g. analysis of variance approach and the generalized estimating equation models [6, 8].

It is worth mentioning that many study designs, other than longitudinal studies, could have continuous repeated measurements and may benefit from the use of linear mixed-effects models. Examples of these study designs are shown in [Supplementary Material, Table S1](#).

## Missing values and correlated variables

In preparation for constructing a linear mixed model, it is important to critically assess the dataset concerning missing values and correlated variables.

**Missing values.** Missing values are quite common in research about cardiothoracic surgery, especially for long-term studies. Dealing with missing values should be done based on the missing mechanisms. Generally, there are 3 mechanisms of missing data: missing completely at random, missing at random and missing not at random (MNAR) [9]. The methods of handling missingness should be selected according to the corresponding mechanism [9]. The proportions of missing values in covariates are presented in [Supplementary Material, Table S2](#). We assumed the missing mechanism of covariates was missing at random and used multiple imputation to get 5 imputed complete datasets. The first and the third of the 5 imputed datasets were used for variable selection and evaluating the reliability of model estimation (sensitivity analysis), respectively.

It is worth noting that missing data is a complicated but highly important part in the construction of linear mixed-effects models, especially when the actual missing mechanism is MNAR. Consulting professional statisticians regarding missing values whenever doubts occur is strongly recommended for clinician researchers. Moreover, researchers should be transparent with their assumptions behind their statistical analyses since plausible assumptions might be necessary for valid estimates.

**Correlated variables.** Many variables play important roles in influencing the outcomes of patients undergoing cardiovascular surgeries, like sex, age, smoking status, etc. Some variables are correlated with each other. In our example dataset, the diameter of homograft is strongly correlated with patients' age, which is quite understandable because children have a smaller RVOT size than adults. Researchers should avoid including variables that show a strong linear correlation in the same model because it could lead to a phenomenon called 'multicollinearity'. Multicollinearity will cause extremely high or low estimates or standard errors in the estimation and influence the 95% confidence interval and *P*-values [10]. Therefore, it is advisable to investigate all the correlations between the covariates by calculating Pearson's or Spearman's correlation coefficients, and one of the highly correlated covariates should be removed. If 2 highly correlated categorical variables are equally clinically meaningful, recoding them into 1 variable could be a solution. Otherwise, clinicians have to weigh their importance or construct models with different options of covariates.

In our example database, 13 variables are available as covariates. Since it is possible that some of them are moderately or highly correlated, their correlation coefficients were calculated. The results are presented in [Supplementary Material, Table S3](#). We set the cutoff value for moderate-to-high correlation to 0.3. In order to avoid multicollinearity, variables with a correlation coefficient higher than 0.3 would be given extra consideration to decide which one to keep. This step of 'variables dropouts' was done by consulting clinicians to ensure the clinically relevant variables would be kept.

Recoding was also done between 2 moderately correlated covariates, 'Number of previous cardiac surgery' and 'first homograft implantation', combining the 2 into a new categorical variable ([Supplementary Material, Table S4](#)) to retain both parts of the information. Considering both age and diagnosis being important covariates, 2 options for a group of covariates were created ([Supplementary Material, Table S5](#)). The procedures of constructing 'model 1' and 'model 2' are almost the same and we only constructed 'model 1' as an example in this article.

## Model specification

When specifying a model, it is essential to take the purpose of constructing a model as a starting point. In this article, the purpose of the model is to clarify the association between time-related RVOT peak gradient patterns and patient characteristics, with the ultimate aim of optimizing clinicians' and patients' decision-making. Clinical experiences and previous evidence should be the basis for selecting potentially associated covariates. A flowchart is presented in [Fig. 1](#) to summarize the general steps of specifying a linear mixed-effects model. To make the abstract concepts more concrete, we applied the example dataset to illustrate these steps.

**Statistical methods in variable selection.** When constructing a model, determining what covariates to include is critical and can be challenging. In certain situations, all available variables can be included as covariates as long as the sample size is big enough (rule of thumb: >10 subjects per covariate in a regression model; more covariates can be included in the longitudinal

setting if repeated measurements are strongly correlated [6]). More proper methods of sample size calculation are also available by using simulations but they are challenging to be widely applied in the clinical research setting [11].

In many practical settings, thorough considerations should be given to select the suitable variables without overfitting. The main idea of doing the variable selection is to evaluate whether a model's performance gets better by adding more variables as covariates. The likelihood ratio test, the Akaike information criterion and the Bayesian information criterion are 3 commonly used methods in evaluating models' performances. The summary of these 3 methods is presented in [Table 2](#). The likelihood ratio test assesses the goodness of fit of 2 nested models based on the ratio of their likelihoods. The best-fit model according to Akaike information criterion/Bayesian information criterion explains the greatest amount of variation using the fewest possible variables [12, 13].

**Interaction.** Interaction describes a condition under which the effect size of 1 specific variable on outcomes is affected by the state of a second variable [14]. Interaction can be present in epidemiology studies, especially when many predictors are considered. Possible interactions should therefore be accounted for when constructing a model. However, it may be difficult or sometimes even impossible to explore all possible interactions, especially in the case of many predictors. Under this circumstance, it is advisable to consider potential interactions from a clinical standpoint.

In longitudinal studies, especially those with a long-term follow-up time, interaction with time is rather important since different groups could have different evolutionary patterns of repeated measurements. For instance, supposing females tend to have slower evolution of RVOT peak gradient than males indicates the interaction between sex and time, and progressive lines of the 2 groups are no longer parallel over time in this scenario ([Supplementary Material, Fig. S1](#)). Sex is an important factor in most clinical settings. Many studies have proven the male-female differences in the field of cardiac surgery [15–17], and therefore, it is interesting to study whether there are male-female differences in the evolution of RVOT peak gradient. The interaction between sex and time was accounted for in our application of the dataset. Other possible interactions were considered after consulting clinicians. All interactions of interest are presented in [Supplementary Material, Table S6](#).

**Nonlinearity.** Assuming a linear relationship between outcome and covariates is not always appropriate in the setting of longitudinal data. We randomly selected 36 patients and plotted their evaluations of repeated measurements, as presented in [Supplementary Material, Fig. S2](#), and obvious nonlinearity between time and repeated measurements was found. In such situations, nonlinearity must be accounted for; otherwise, the model is invalid. Usually, multiple measurements are required for nonlinearity consideration, and at least 3 measurements on average are needed to account for nonlinearity between covariates and outcome.

There are several ways of addressing nonlinear relations: transformation (like logarithmic transformation), splines and polynomials (regular or fractional) [18, 19]. Transformations change all

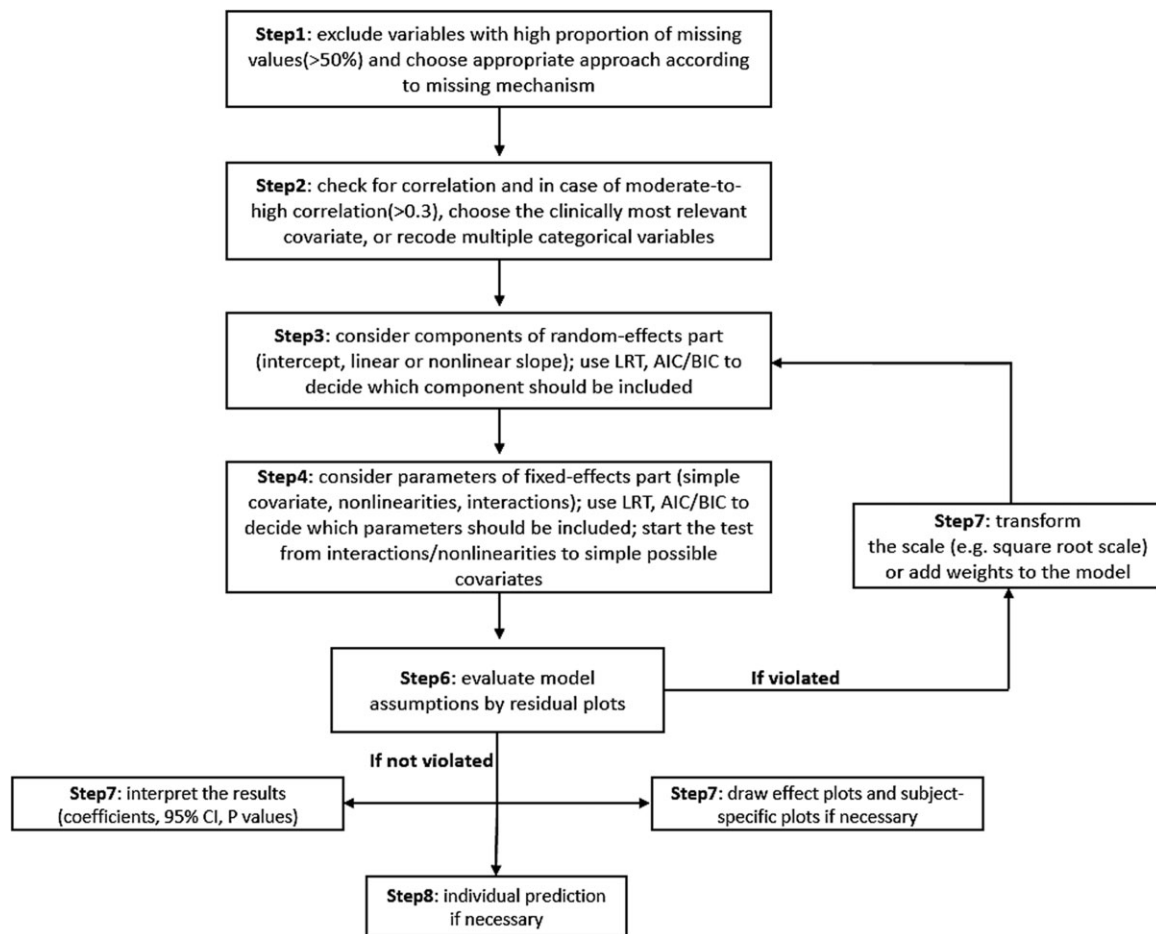


Figure 1: Flowchart of linear mixed-effects models construction.

Table 2: Explanations of different methods in evaluating models

	Mechanism	Indications	Notes
AIC/BIC	Penalized-likelihood criteria: rewarding the goodness of fit (likelihood function) and penalizing the increased number of parameters (overfitting); model with lower AIC/BIC is superior to the one with higher values; compared to AIC, BIC penalizes more for overfitting [12, 13]	Non-nested/nested models  Not the first choice if the models are nested	No certain cutoff value of AIC/BIC difference for selecting the superior model, better to check both  If the 2 contradicts with each other, AIC tends to select more elaborate models than BIC since the latter penalizes more heavily for complexity of the model
LRT	Assessing the goodness of fit of 2 nested models based on the ratio of their likelihoods	Nested models Not applicable to non-nested models	The LRT performs less efficiently in testing optimal random-effects part because of being on a boundary condition

AIC: Akaike information criterion; BIC: Bayesian information criterion; LRT: likelihood ratio test.

the relationships between covariates and response variables. However, nonlinearity might only exist between a part of covariates and response variables. Hence, transformation is not flexible in dealing with the 'partial' nonlinearities compared to splines and polynomials. Splines partition the curve into several small parts by placing internal knots and are more flexible than polynomials in accounting for nonlinearity [20, 21]. Among different types of splines, natural (cubic) splines are less erratic at the boundaries of the data and are preferable in capturing

nonlinearity in the mixed-effects models [20]. The number and position of the internal knots should be selected when applying natural cubic splines. However, it is uneasy to choose the optimal internal knots that can capture the nonlinearity sufficiently without overfitting. It is recommended to determine knot selection with a biostatistician. Natural cubic splines with 2 internal knots were used in our application and procedures of our knots selection are presented in [Supplementary Material, Table S7](#) and [Supplementary Material, Figs S3–S7](#).

**Procedures in model construction.** The model specification should start from the random-effects part, then move to the fixed-effects part. At the time of specification of the random-effects part, all possible predictors should be put in the fixed-effects part, including interactions and nonlinearity. The components of the random-effects part and fixed-effects part should be tested in the order of 'from complex to simple'. Plotting individual evolutions over time could help clinicians visualize 'time-nonlinearity' and select the optimal number of internal knots. The scale of outcomes in our application is the square root of RVOT peak gradient because the original scale violates the 'variance constant' assumption, which means the residuals have constant variance at every level of predicted outcomes. Further explanations of this violation will be given in 'model assumptions' part.

**Random-effects part** Intra-correlations between repeated measurements of the same patient are very important for the modeling of evolution patterns. One important element to illustrate the intra-correlations is the variance-covariance matrix. In linear mixed-effects models, this matrix is incorporated into the random-effects part. More explanations of this matrix have been presented in [Supplementary Material, Text S1](#).

To incorporate the appropriate structure of the variance-covariance matrix in random-effects part, careful considerations of components that should be included in the random-effects part are essential. Some publications have introduced more details about the variance-covariance matrix [22, 23]. In general, it is advisable not to have the very complex structure of random-effects and correlation matrix.

Several available components, including random intercepts and linear/nonlinear slopes, should be considered in the random effects. These components are visualized in Fig. 2. In Fig. 2A, different patients had different baseline values (random intercept); in Fig. 2B, patients did not only have different baseline values but also showed different evolutions (random intercept and random slope); in Fig. 2C, nonlinear slopes were observed in each participant (random intercept and nonlinear slope).

Sixteen patients in the example dataset were selected to display their individual evolutions (Fig. 3). Different baseline values and nonlinear slopes were observed among these patients. Hence, an elaborate random-effect structure including random intercepts and

non-linear random slopes was chosen. The most complex structure was fitted firstly and compared with parsimonious structures. The results are shown in [Supplementary Material, Table S7](#). In the case of many data points, statistical tests may favour more complex structures of random effects but are modelling the noise (i.e. overfitting) in fact. Therefore, splines with various internal knots are visualized in [Supplementary Material, Figs S3–S5](#). Nonlinearity was not sufficiently captured in [Supplementary Material, Fig. S3](#) while no big differences were observed between [Supplementary Material, Figs S4 and S5](#). Thus, splines with 2 internal knots ([Supplementary Material, Fig. S4](#)) were sufficient to capture nonlinearity in the random-effects part.

**Fixed-effects part** After finishing specifying the random-effects structure, the next step is simplifying the elaborate fixed-effects part. In this case, one should also balance the statistical importance versus the clinical importance. Several statistically non-significant variables could be kept if they are of high clinical relevance. It is worth noting that eliminating a non-significant but clinically relevant variable (i.e. sex or age) might lead to biased estimates of other variables [24]. Some stepwise variable-selection methods are only applicable to the dataset with a very large sample size [24]. Therefore, only eliminations of interactions and nonlinearities were tested statistically to simplify the fixed-effects part of the model. The results are shown in [Supplementary Material, Tables S7 and S8](#). Selecting internal knots of natural cubic splines in the fixed-effects part was done and 2 internal knots were applied finally. The procedures of knot selection are displayed in [Supplementary Material, Table S7](#) and [Supplementary Material, Figs S6 and S7](#).

**Results' interpretations** The estimated coefficients of the final model are shown in [Supplementary Material, Table S9](#). The estimated coefficient for age was -0.015, which implies a 1-year increase in age at operation causes a 0.015-mmHg decrease of the square root of RVOT peak gradient on average, given all other covariates are fixed. The coefficient for male sex was 0.582. This indicates male patients tend to have a 0.582-mmHg higher square root of RVOT peak gradient than female patients with equal values of other covariates.

Effect plots are useful to give a clear illustration of the results, especially when nonlinearities and interactions exist. These plots

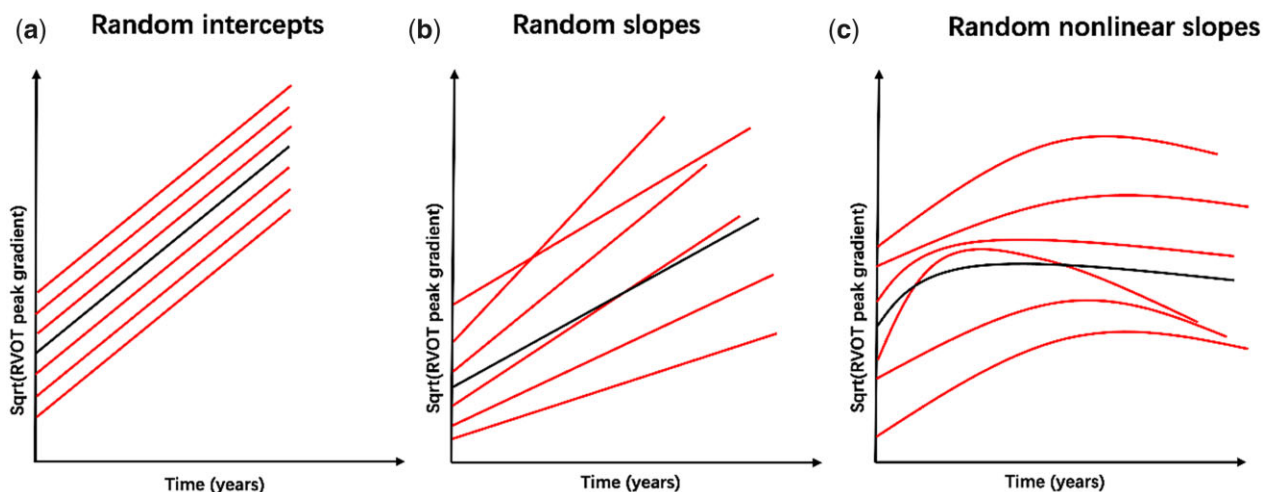
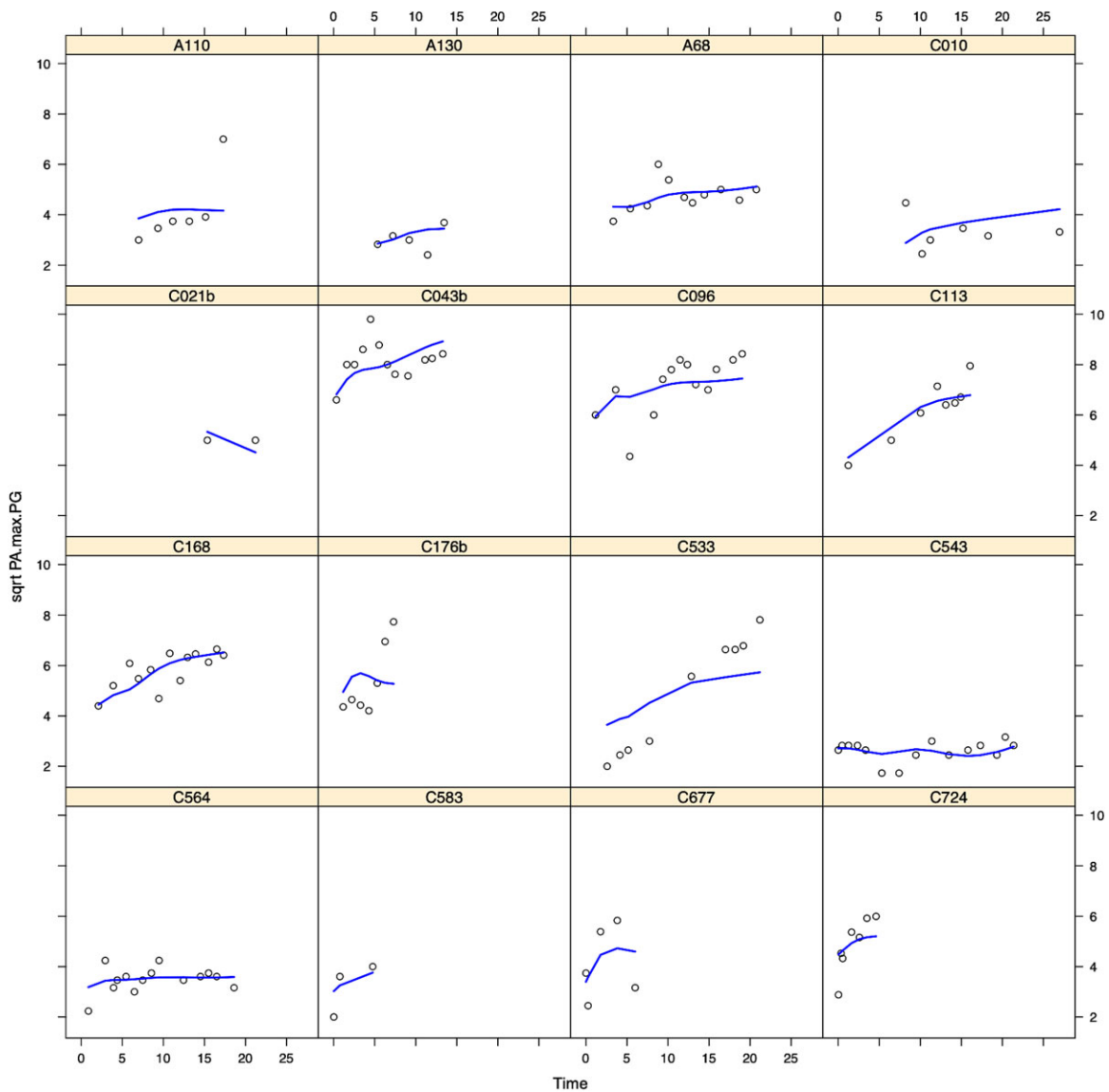


Figure 2: Explanations of random intercepts, random slopes and random nonlinear slopes.



**Figure 3:** Sixteen patients' evolutions of square root of right ventricular outflow tract peak gradient showed nonlinearity.

are figures depicting the predicted values of the outcome for specific combinations of covariates' values and are helpful in visualizing the predicted results at a population level. To create effect plots, values of covariates need to be specified, e.g. female, 21-year-old. The effect plots of the 21-year-old male and female are presented in Fig. 4. The evolutions of the square root of RVOT peak gradient of the 2 groups of patients indicate males and females have different progressive courses concerning the square root of RVOT peak gradient.

Besides effect plots that reflect average evolution, subject-specific plots show curves of individual predicted values over time. Providing prediction on an individual level is one of the strengths of linear mixed-effects models. The marginal predictions (illustrating predictions on average level) and subject predictions for 16 selected patients are displayed in [Supplementary Material, Fig. S8](#). The obvious superiority of subject predictions over marginal predictions was observed. It highlights the advantages of linear mixed-effects models in predictions on the individual level. To illustrate the individualized predicted values, we

used patients that were included in the model. Individualized dynamic predictions could also be obtained for new patients by using linear mixed-effects models. However, this is out of the scope of this manuscript and only brief introduction will be given in the discussion of 'extensions of mixed-effects models to accommodate survival data: joint models and dynamic prediction'.

### Model assumptions

There are several assumptions of the linear mixed-effects model. The linearity assumption is one of them. It assumes linearity of the relationship between predictors and response variables. Specific patterns in the plot of residuals versus predicted response variable indicate potential violation of this assumption, as presented in [Supplementary Material, Fig. S9](#) (in this model, nonlinearity between time and square root of peak RVOT gradient was not accounted). To address this violation, nonlinear relations could be incorporated, as discussed previously. Besides, the model requires homoscedasticity (constant variance) and normal

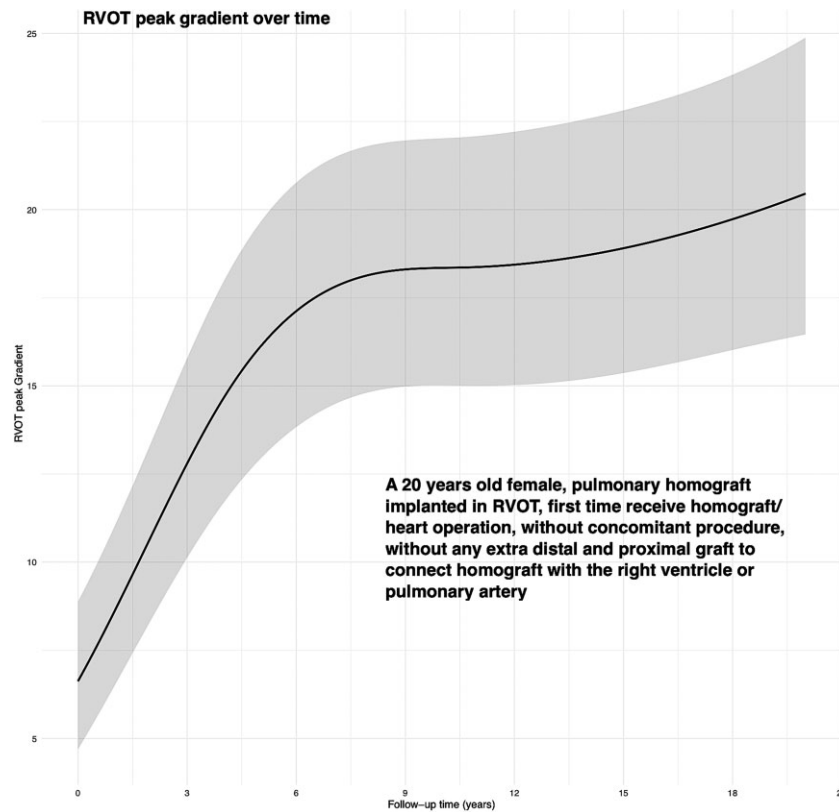


Figure 4: Effect plots of the finally fitted model, with considerations of interactions and nonlinearities.

residual assumptions. The residual is the difference between the observed value and the predicted value derived from the fitted model. If the model specification is valid, its residuals assume homoscedasticity and follow the normal distribution with a mean of zero [25]. All the residuals should be randomly scattered without any obvious tendency. Otherwise, these assumptions are violated.

Residual plots, including Q-Q plots and fitted values versus residuals plots, are usually illustrated to check the distribution of residuals. Ideally, residuals should be located completely along the diagonal in the Q-Q plot and distributed evenly, without any patterns in fitted values versus residuals plots. Besides, the average line of residuals should be almost or entirely equal to zero. It is extremely difficult to get perfect residual plots, because specifying a model without any bias is almost impossible. Thus, a slight discrepancy from normality is acceptable. If an obvious discrepancy appears, we can address it by transforming the scale of the outcome (in case of a continuous outcome, e.g. square root scale) or adding weights to the model [8].

In our application, the original scale of serial peak gradient was used initially, and inconstant variance was observed in residual plots (Supplementary Material, Fig. S10). After transforming to square root scale, the residuals scattered more evenly (Supplementary Material, Fig. S10). To test whether the residuals fulfill the normality assumption, the Q-Q plots (Supplementary Material, Fig. S11) and fitted values versus residuals plots (Supplementary Material, Fig. S12) were illustrated, showing no systematic deviations and approximately normally distributed residuals. In fitted values versus residuals plots, the conditional residuals were scattered almost evenly around the centreline.

## Comparison with linear regression model

To illustrate that a linear regression model may cause biased estimates compared to the linear mixed-effects model, a linear regression model that ignores the correlations between repeated measurements within the same patient was constructed. The results are shown in Supplementary Material, Table S9 and Supplementary Material, Fig. S13. Major differences were observed between the 2 models on estimates and individual prediction. Compared to linear regression models, linear mixed-effects models can provide more precise subject-specific predicted values.

## DISCUSSION

Nowadays, many patients undergoing cardiothoracic surgery have a long-term follow-up after their surgery. Important biomarkers, such as echocardiographic measurements and N-terminal pro b-type natriuretic peptides, are measured repeatedly over time and assist clinicians in clinical decision-making. Considering the evolution over time of these measurements rather than only considering a 'snapshot' of 1 measurement can be of great added value in clinical decision-making. Therefore, the requirement of an adequate statistical method that can capture the evolutions of repeated measurements is increasingly necessary for the research of cardiothoracic surgery. Linear mixed-effects models are applicable to continuous longitudinal data as well as other grouped continuous data (multilevel data) and flexible in dealing with unbalanced measurements, a very common



situation in the follow-up after cardiac surgery. In this article, we provided step-by-step guidance in using mixed-effects models with a dataset application. Several aspects were discussed, including correlated covariates, interactions and nonlinearities. In particular, as nonlinearity may exist in both parts of fixed- and random-effects part, statistical testing combined with visualization is highly recommended to deal with it. It is worth noting that the interactions between sex and other covariates (e.g. sex and treatment procedures) are usually vital to study in cardiothoracic research as sex (and gender) is a strong determinant of cardiovascular outcomes.

Interpreting estimates of coefficients from a model with multiple interactions and nonlinearities is not straightforward. Given the fact that interactions and nonlinearities are common in cardiothoracic research, presenting the results of the model with effect plots is especially encouraged in the setting of cardiothoracic surgery. Effect plots are very intuitive and can be easily used to visualize evolutions of repeated measurements. As with every statistical analysis, the assumptions of the linear mixed-effects models should be evaluated after model construction. The 'Q-Q plot' and 'fitted-values-versus-residuals plot' are 2 common ways of evaluating assumptions. No systematic trend should be observed if the assumptions are valid.

The R syntax used in the analysis of our example data set is provided in [Supplementary Material 2](#).

## Comparison with simplified approaches of longitudinal data analysis

The commonly used simplified approaches to longitudinal data analysis are simple linear regression models and time-to-event analyses. Compared with linear mixed-effects models, simple linear regression models ignore the intra-correlations of repeated measurements, which could cause bias to the estimates. To provide more valid results, researchers should opt for mixed-effects model when they model repeated measurements.

Apart from the linear regression model, the time-to-event analysis is another widely used method in dealing with longitudinal data, by transforming these measurements into events according to the definitions. However, this methodology causes loss of information and may lead to spurious conclusions in the setting of dynamic outcomes, such as valve function, because survival models assume events occur at an instant in time. However, many outcomes of importance are conditions or processes that evolve with time, such as return of regurgitation after valve repair. Moreover, values for these outcomes are captured at discrete instances in time ('snapshots'). Snapshots are subject to many biases, especially when a condition changes rapidly but snapshots are taken infrequently. If follow-up occurs only when symptoms recur, the prevalence of undesirable change may be overestimated [4]. In addition, an event is not reversible, which means once an event occurs, it is impossible to go back to the 'no event' status unless the subject with the subsequent event is treated as a new one. Thus, 1 patient can be attributed to 'event' by an inaccurate measurement and stay in 'event' status forever. It could cause bias, even wrong conclusions if too many patients have 'inaccurate measurements'. The 2008 Akins guidelines [4] for reporting mortality and morbidity after cardiac valve interventions recommend the use of longitudinal data analysis to assess valve function over time rather than survival analysis.

## Extensions of mixed-effects models

**Joint models.** Researchers might want to investigate whether a repeatedly measured variable, continuous or categorical, is a risk factor for an event. For instance, with the dataset used in this article, clinicians might want to explore the relation between the evolution of the square root of RVOT peak gradient and the hazard of death for patients receiving homograft in the pulmonary position. In this situation, a framework connecting repeated measurements with survival status can be used. This framework is called joint models of longitudinal and survival data [26]. One can also use this framework to account for informative censoring in longitudinal data (MNAR). Its application in heart valve function has been introduced by Andrinopoulou *et al.* [27].

**Dynamic prediction.** Nowadays, as disease populations, clinical practice and healthcare systems are constantly evolving, some predictive models are quickly becoming outdated and less accurate over time. A predictive model capable of 'self-updating' by including new measurements for a specific subject to calibrate the prediction is necessary. A methodology called dynamic prediction can achieve the purpose, for both mixed-effects models and joint models [28]. This statistical method is becoming increasingly helpful, and more personalized and dynamic predictions can be used in clinical decision-making for each patient by using dynamic prediction.

## CONCLUSION

Linear mixed-effects models can be used in modelling repeated measurements to account for their evolutions over time and give more valid estimates compared to linear regression models in the setting of cardio-thoracic surgery outcomes research.

## SUPPLEMENTARY MATERIAL

[Supplementary material](#) is available at *EJCTS* online.

**Conflict of interest:** All authors have nothing to disclose.

## Data Availability Statement

The data underlying this article will be shared upon reasonable request to the corresponding author.

## Author contributions

**Xu Wang:** Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Visualization; Writing—original draft.  
**Eleni-Rosalina Andrinopoulou:** Conceptualization; Investigation; Methodology; Resources; Validation; Writing—review & editing.  
**Kevin M. Veen:** Conceptualization; Methodology; Validation; Writing—review & editing.  
**Ad J.J.C. Bogers:** Project administration; Validation; Writing—review & editing.  
**Johanna J.M. Takkenberg:** Conceptualization; Methodology; Supervision; Validation; Writing—review & editing.

## Reviewer information

European Journal of Cardio-Thoracic Surgery thanks the anonymous reviewer(s) for their contribution to the peer review process of this article.

## REFERENCES

- [1] Cuypers JA, Menting ME, Konings EE, Opić P, Utens EM, Helbing WA *et al* Unnatural history of tetralogy of Fallot: prospective follow-up of 40 years after surgical correction. *Circulation* 2014;130:1944–53.
- [2] van der Linde D, Konings EE, Slager MA, Witsenburg M, Helbing WA, Takkenberg JJ *et al* Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. *J Am Coll Cardiol* 2011;58:2241–7.
- [3] Otto CM, Nishimura RA, Bonow RO, Carabello BA, Erwin JP 3rd, Gentile F *et al* 2020 ACC/AHA guideline for the management of patients with valvular heart disease: executive summary: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* 2021;143:e35–e71.
- [4] Akins CW, Miller DC, Turina MI, Kouchoukos NT, Blackstone EH, Grunkemeier GL *et al*; EACTS. Guidelines for reporting mortality and morbidity after cardiac valve interventions. *Ann Thorac Surg* 2008;85:1490–5.
- [5] Marill KA. Advanced statistics: linear regression, part I: simple linear regression. *Acad Emerg Med* 2004;11:87–93.
- [6] Verbeke G, Molenberghs G. *Linear Mixed Models For Longitudinal Data*. New York: Springer, 2005.
- [7] Chen PC, Sager MS, Zurakowski D, Pigula FA, Baird CW, Mayer JE Jr *et al* Younger age and valve oversizing are predictors of structural valve deterioration after pulmonary valve replacement in patients with tetralogy of Fallot. *J Thorac Cardiovasc Surg* 2012;143:352–60.
- [8] Hedeker D, Gibbons RD. *Longitudinal Data Analysis*. Hoboken, NJ/Chichester: Wiley; John Wiley distributor, 2006.
- [9] Papageorgiou G, Grant SW, Takkenberg JJM, Mokhles MM. Statistical primer: how to deal with missing data in scientific research? *Interact CardioVasc Thorac Surg* 2018;27:153–8.
- [10] Greenland S, Daniel R, Pearce N. Outcome modelling strategies in epidemiology: traditional methods and basic alternatives. *Int J Epidemiol* 2016;45:565–75.
- [11] Green P, MacLeod CJ. SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol Evol* 2016;7:493–8.
- [12] Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr* 1974;19:716–23.
- [13] Schwarz G. Estimating the dimension of a model. *Ann Statist* 1978;6:461–4.
- [14] Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Cham, Switzerland: Springer, 2019. <https://doi.org/10.1007/978-3-030-16399-0>.
- [15] Mokhles MM, Soloukey Tbalvandany S, Siregar S, Versteegh MIM, Noyez L, van Putte B *et al* Male-female differences in aortic valve and combined aortic valve/coronary surgery: a national cohort study in the Netherlands. *Open Heart* 2018;5:e000868.
- [16] Romeo JLR, Mokhles MM, van de Woestijne P, de Jong P, van den Bosch A, van Beynum IM *et al* Long-term clinical outcome and echocardiographic function of homografts in the right ventricular outflow tract. *Eur J Cardiothorac Surg* 2019;55:518–26.
- [17] Mokhles MM, Siregar S, Versteegh MI, Noyez L, van Putte B, Vonk AB *et al* Male-female differences and survival in patients undergoing isolated mitral valve surgery: a nationwide cohort study in the Netherlands. *Eur J Cardiothorac Surg* 2016;50:482–7.
- [18] Veen KM, de Angst IB, Mokhles MM, Westgeest HM, Kuppen M, Groot CAU *et al* A clinician's guide for developing a prediction model: a case study using real-world data of patients with castration-resistant prostate cancer. *J Cancer Res Clin Oncol* 2020;146:2067–75.
- [19] Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *J R Stat Soc Ser C Appl Stat* 1994;43:429–67.
- [20] Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M. A review of spline function procedures in R. *BMC Med Res Methodol* 2019;19:46.
- [21] Binder H, Sauerbrei W, Royston P. Comparison between splines and fractional polynomials for multivariable model building with continuous covariates: a simulation study with continuous response. *Stat Med* 2013;32:2262–77.
- [22] Galecki ATBT. *Linear Mixed-Effects Models Using R: A Step-by-Step Approach*. New York, NY: Springer, 2013.
- [23] Verbeke GMG. *Linear Mixed Models for Longitudinal Data*. New York: Springer, 2006.
- [24] Heinze G, Dunkler D. Five myths about variable selection. *Transpl Int* 2017;30:6–10.
- [25] Schielzeth H, Dingemanse NJ, Nakagawa S, Westneat DF, Allogue H, Teplitsky C *et al* Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods Ecol Evol* 2020;11:1141–52.
- [26] Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data*. Boca Raton, Florida: CRC Press, 2012.
- [27] Andrinopoulou ER, Rizopoulos D, Jin R, Bogers AJ, Lesaffre E, Takkenberg JJ. An introduction to mixed models and joint modeling: analysis of valve function over time. *Ann Thorac Surg* 2012;93:1765–72.
- [28] Andrinopoulou ER, Rizopoulos D, Geleijnse ML, Lesaffre E, Bogers AJ, Takkenberg JJ. Dynamic prediction of outcome for patients with severe aortic stenosis: application of joint models for longitudinal and time-to-event data. *BMC Cardiovasc Disord* 2015;15:28.