An Overview of Methods for Interval-Censored Data with an Emphasis on Applications in Dentistry

Emmanuel Lesaffre[1], Arnošt Komárek[1]

and Dominique Declerck[2]

[1] Katholieke Universiteit Leuven, Biostatistical Centre, Kapucijnenvoer 35, B–3000, Leuven, Belgium

[2] Katholieke Universiteit Leuven, School of Dentistry, Kapucijnenvoer 7, B–3000, Leuven, Belgium

Corresponding author: Emmanuel Lesaffre

E-mail: emmanuel.lesaffre@med.kuleuven.ac.be

Tel.: +32 - 16 - 336896

Fax: +32 - 16 - 336900

ABSTRACT. Interval-censored time-to-event data occur in many medical areas with dentistry or AIDS research being typical representatives. This article reviews methods for the analysis of such data with an emphasis on the use of the accelerated failure time (AFT) model. A flexible AFT model (avoiding parametric assumptions on the distribution of the error term) is described more in detail and used to solve a typical dental question in a longitudinal oral health study.

KEY WORDS. Accelerated Failure Time Model; Interval-Censored Data; Regression Model.

---

1. Introduction

Standard survival methods assume that individuals are followed over time for the occurrence of a specific event. The time to the event is referred to as duration time. If at the end of the observation period the event has not been observed, the time to the event is called "right-censored". However, in dentistry (and other areas of medical research), the occurrence of the event of interest can often be recorded only at planned (or unplanned) visits which gives rise to "interval-censored data". A typical example is the time to caries development or to emergence of a tooth. Indeed, in case of a cavity or of emergence the event is often observed after some delay, say at planned (or even unplanned) visits.

For right-censored data, a battery of statistical tests and techniques is available to tackle most research questions under a variety of statistical assumptions. Also, commercial software is available for many procedures. For interval-censored data, statistical techniques are less well developed and their statistical properties are much more complex. But above all, almost no (commercial) software is available.

In this paper we first review the current statistical methods on interval-censored data. Then we describe a recently developed survival method which is useful for interval-censored data and operates under mild statistical assumptions.

To proceed, some notation is needed. Let T_i $(i = 1, \ldots, n)$ be the random variable recording the duration time of the ith individual in the sample. With interval-censored data, instead of T_i we only observe intervals $[L_i, R_i]$, where $L_i \leq T_i \leq R_i$. Note that this does not rule out exactly observed, right-censored and left-censored data for which $L_i = R_i = T_i$, $R_i = \infty$ and $L_i = 0$, respectively. Often, additionally a vector $\boldsymbol{x}_i$ of covariates $(i = 1, \ldots, n)$ is recorded. A typical question is then whether the distribution of T_i depends on the covariates.

To illustrate our analyses, we consider the Signal Tandmobiel® study. It is a longitudinal oral health study performed in Flanders in 1996–2001 involving 4 468 schoolchildren (2 315 boys and 2 153 girls) born in 1989 who were annually examined. More on the design of the study can be found in Vanobbergen et al.[1] In this paper, we focus on the distribution of emergence times of permanent maxillary right premolars (teeth 14 and 15 in European dental notation). Adequate knowledge of timing and patterns of tooth emergence is useful for diagnosis and treatment planning in paediatric dentistry and orthodontics. It is anticipated, that the distribution of emergence times of a particular tooth is different for boys and girls. For that reason, we used the covariate *gender* (0 for boys and 1 for girls) in our models. Additionally, it was of dental interest to check whether the distribution of the emergence time of a permanent tooth changes when the primary predecessor of the permanent experienced caries or not. For this, we included a binarised *dmf* score pertaining to the predecessor as a covariate, $dmf = 1$ if the primary predecessor of that permanent tooth was recorded as decayed, or missing due to caries, or filled and 0 otherwise. Due to the design of the study (annual planned examinations), the response variable – time to emergence of a particular tooth – is interval-censored with intervals of length equal to approximately 1 year.

For each tooth separately, the following typical goals in survival analysis are of interest: (1) Estimate the survival function $S(t|\boldsymbol{x}) = P(T > t|\boldsymbol{x})$ or equivalently the cumulative distribution function (cdf) $F(t|\boldsymbol{x}) = 1 - S(t|\boldsymbol{x})$. In the context of tooth emergence the cdf is known as *the emergence curve*; (2) Compare the distributions of emergence times of two sub-populations specified by two different covariate vectors (e.g. compare children with sound and decayed primary teeth, or boys and girls); (3) Set up a regression model to describe the distribution of emergence times for populations with different values of covariates.

In Section 2, we review some methods that address the above questions in the context of interval censoring. In this aspect we focus on the accelerated failure time (AFT) model which is the competitor of the more widely used proportional hazards (PH) model. The third section briefly describes the flexible AFT model of Komárek, Lesaffre and Hilton[2] which can be used for analysis of interval-censored data without making strong distributional assumptions. In Section 4, the use of this model will be illustrated using the Signal Tandmobiel® data. In the final section future research directions are indicated.

2. Review of methods for interval-censored data

A variety of methods (non-, semi- and fully parametric) for right-censored data have been developed. Further, commercial software is available to support these techniques. In contrast, for interval-censored data commercial software seems to be available only for non-parametric estimation of survivor curves and for parametric modelling, besides of course the user-written programs. Further, until recently only few methods were available. That is why, in practice, modelling with interval-censored data is often mimicked by methods developed for right-censored data. For this, the interval needs to be replaced by an exact time. The most common assumption is that the event occurred at the midpoint of the interval. However, applying methods for right-censored data on these artificial fixed points can lead to biased and misleading results, see e.g., Rücker and Messerer[3]; Law and Brookmeyer[4]; Odell et al.[5]; and Dorey et al.[6] In this paper, we will review appropriate methods to deal with interval-censored data and link them to the corresponding (classical) method for right-censored data.

2.1. **Estimation of survival/cumulative distribution function.**

The classical non-parametric method is given by Kaplan and Meier[7]. Its interval-censored data counterpart was first proposed by Peto[8]. Turnbull[9] improved the numerical algorithm to estimate the

survival function using his so called *iterative self-consistency* algorithm, which is, in fact, an EM type (Dempster, Laird and Rubin[10]) of algorithm. Nowadays, it is implemented in the S-PLUS function `kaplanMeier`.

A valuable alternative to non-parametric procedures is obtained by smoothing the survival or equivalently the density function or the hazard function. In situations where it can be assumed that the event-times are continuously distributed, we even get more realistic, not step-wise, estimates. One such method, applicable directly to both right- and interval-censored data is given by Kooperberg and Stone[11] who smooth the density using splines and also provide software in the form of the R[12] package `logspline` downloadable from `www.R-project.org` or the S-PLUS library `splinelib` downloadable from *StatLib*. Splines in the smoothing of the hazard function are then exploited by the approach of Rosenberg[13].

2.2. **Comparison of two survival distributions.** For right-censored data, many non-parametric tests for comparing two survival curves are available, e.g. the log-rank test (Mantel[14]), the Gehan generalization of the Wilcoxon test (Gehan[15]), the Peto-Prentice generalization of the Wilcoxon test (Peto and Peto[16]; Prentice[17]) and the weighted Kaplan-Meier statistic (Pepe and Fleming[18]).

The Gehan-Wilcoxon test has been adopted to interval-censored data by Mantel[19], while the interval-censored version of the Peto-Prentice-Wilcoxon test is presented by Self and Grossmann[20]. The log-rank test for interval-censored data is given by Finkelstein[21]. Further, Petroni and Wolfe[22] discuss the weighted Kaplan-Meier statistic in the context of interval-censoring. Finally, Fay[23] derives a general class of linear-rank tests for interval-censored data which covers, as special cases, the Wilcoxon-based tests.

Regrettably, the asymptotic properties of the above methods assume *the grouped continuous model*, which implies that the status of each subject is checked at the same timepoints (in the study time scale) whose number is fixed or that observed intervals are grouped in

such a way. For our dental application, this would mean that the emergence status of the teeth was checked at prespecified ages, the same for all children. Obviously, such setting is too restrictive in many practical situations. E.g., in our case, a particular child was checked by a dentist-examiner on a prespecified day of the year, irrespective of his or her age.

The grouped continuous model assumption is necessary to be able to apply the standard maximum likelihood theory to interval-censored data measured on a continuous scale without making any parametric assumptions. Only recently, Fang, Sun and Lee[24] developed a test statistic, based on the weighted Kaplan-Meier statistic of Pepe and Fleming[18] that does not require the grouped continuous model assumption. Finally, Pan[25] offers two-sample test procedures obtained by combining standard right-censored tests and multiple imputation that allows, in contrast to single (e.g. midpoint) imputation mentioned at the beginning of this section, to draw appropriately the statistical inference.

Unfortunately, none of the above described approaches is available in any commercial statistical package.

2.3. **Proportional hazards regression model.** The proportional hazards (PH) model (Cox[26]) is the most popular regression model for right-censored data. For a given covariate vector $\boldsymbol{x}$, the hazard function $h(t|\boldsymbol{x}) = \lim_{dt \to 0_+} P(t \leq T < t + dt | T \geq t, \boldsymbol{x})/dt$ is expressed as the product of an unspecified baseline hazard function $h_0(t)$ and the exponential of a linear function of the covariates, i.e.

$$h(t|\boldsymbol{x}) = h_0(t) \ \exp(\boldsymbol{\beta}'\boldsymbol{x}). \tag{1}$$

The regression parameter vector $\boldsymbol{\beta}$ is estimated by maximizing a partial likelihood (Cox[27]) which treats h_0 as nuisance and does not estimate it. However, when the baseline hazard h_0 is of interest as well, e.g. for prediction purposes, its non-parametric estimate can be obtained using the method of Breslow[28].

To extend the PH model to interval-censored data, basically four types of approaches can be found in the literature. Firstly, the baseline hazard h_0 can be parametrically specified and standard maximum likelihood theory applied to estimate the regression parameters. However, the parametric assumptions can cause bias in inference if incorrectly specified.

A second class of methods makes use of a combination of multiple imputation and methods for right-censored data (Satten[29]; Satten, Data, Williamson[30]; Goggins et al.[31]; Pan[32]). A disadvantage of these methods is, however, that they are highly computationally demanding and the fact that the procedures they use to impute missing data have a relatively *ad hoc* nature.

A third approach, suggested by Finkelstein[21] and Goetghebeur and Ryan[33] resembles most the method of Cox[26] combined with that of Breslow[28]. Indeed, in both papers the baseline hazard h_0 is estimated non-parametrically on top of estimating the regression coefficients. Whereas the method of Finkelstein relies on the grouped data assumption, Goetghebeur and Ryan developed an EM-type procedure that relaxes that assumption. Moreover, the second approach seems to be the only one that reduces to a standard Cox model when interval-censoring reduces to right-censoring ($R_i = \infty$).

Finally, methods that smoothly estimate h_0 are a trade-off between parametric modelling that allows for a straightforward maximum likelihood estimation of the parameters and semi-parametric models with a completely unspecified baseline hazard h_0. Kooperberg and Clarkson[34] suggest regression splines, while Betensky et al.[35] use local likelihood smoothing to model the baseline hazard. A nice feature of these methods is that predictive survival and hazard curves are directly available and moreover, they are smooth rather than step-wise as in the case of semi-parametric estimation. The software for the approach of Kooperberg and Clarkson[34] is included in the previously mentioned R package `logspline` or S-PLUS library `splinelib`.

2.4. **Accelerated failure time model.** A useful, however less frequently used alternative to the PH model is the accelerated failure time (AFT) model. In this case, the effect of a covariate is an acceleration or deceleration of the event time. For a vector of covariates $\boldsymbol{x}$ the effect is expressed by the parameter vector $\boldsymbol{\beta}$ in the following way:

$$T = \exp(\boldsymbol{\beta}'\boldsymbol{x})\tau, \tag{2}$$

where τ is a baseline survival time. On the logarithmic scale, this model becomes a simple linear regression model

$$\log(T) = \boldsymbol{\beta}'\boldsymbol{x} + \varepsilon, \tag{3}$$

with $\varepsilon = \log(\tau)$. Usually one assumes that the random variable ε has a density $f^*(e)$ from the location-scale family, i.e. $f^*(e) = \sigma^{-1} f\{\sigma^{-1}(e-\alpha)\}$, where $f(\cdot)$ has location parameter $= 0$ and scale parameter $= 1$, α is a location parameter and σ a scale parameter.

A parametric AFT model assumes that $f(e)$ is a density of a specific type (e.g. normal, logistic or Gumbel). In that case, the parameters α, σ and $\boldsymbol{\beta}$ can be estimated easily using maximum likelihood techniques even with interval-censored data, see, e.g., `survreg` in R, `SurvReg` in S-PLUS or `lifereg` in SAS. Evidently, the parametric assumptions affect the shape and character of the resultant survival or hazard curves which, in the case of an incorrect specification, is undesirable, especially when prediction is of interest.

On the other hand, semi-parametric procedures for the AFT model leave the density $f^*(e)$ unspecified and provide only the estimate of the regression parameter vector $\boldsymbol{\beta}$. The whole class of these methods for right-censored data, based on linear-rank tests is described comprehensively by Kalbfleisch and Prentice[36], Chapter 7. Their extension to interval-censoring was studied by Rabinowitz, Tsiatis and Aragon[37] and Betensky, Rabinowitz and Tsiatis[38]. Though, both approaches are practically applicable only with low-dimensional covariate vectors $\boldsymbol{x}$. Furthermore, in contrast to the PH model where for right-censored data

the baseline hazard can be estimated non-parametrically (Breslow[28]) no similar approach is available for the AFT model which implies that semi-parametric procedures cannot be used when prediction is of interest.

Approaches based either on multiple imputation (Pan and Louis[39]; Pan and Connett[40]) or its combination with smoothing (Pan and Kooperberg[41]) constitute more promising alternatives. The methods of the mentioned papers are applied in the context of multivariate survival data, however their application to univariate data is straightforward.

Recently, Komárek et al.[2] suggested and implemented a maximum-likelihood based approach for the AFT model that exploits penalized smoothing of the baseline density $f^*(e)$. As such, the method offers both estimates of regression parameters and baseline density without making any strong parametric assumptions. As a consequence, realistic estimates of survival or hazard curves can easily be obtained. In Section 3, we describe this method briefly and in Section 4 the method is applied to the Signal Tandmobiel® data.

2.5. **Accelerated failure time model versus proportional hazards model.** Both the PH as well as the AFT model make an explicit assumption about the effect of covariates on the hazard function. The effect of covariates on the hazard function in the PH model is given by (1). For the AFT model, the hazard function of the random variable $T = \exp(\boldsymbol{\beta}'\boldsymbol{x})\tau$ is

$$h(t \mid \boldsymbol{x}) = \exp(-\boldsymbol{\beta}'\boldsymbol{x}) h_0\big\{\exp(-\boldsymbol{\beta}'\boldsymbol{x})\ t\big\}. \tag{4}$$

The assumed different effect of a covariate on the baseline hazard for the PH and AFT model is exemplified in Figure 1. It is seen that in the AFT model the effect of covariates on the baseline hazard function is multiplicative, like in the PH model, and on top of that an acceleration or deceleration of the time scale. Secondly, in the AFT model the hazard is increased for $\beta < 0$ whereas in the PH model for $\beta > 0$.

Further, it is generally true that it is not always possible (e.g. due to lack of knowledge) to include all relevant covariates in the model. One of the advantages of the AFT model is that the regression parameters of the included covariates do not change when other, important, covariates are omitted. Of course, the neglected covariates have an impact on the distribution of the error term ε which is typically changed into one with larger variability. Such change, however, is of no importance (except that it influences the precision with which the regression parameters of the included covariates are estimated) when semi-parametric or smoothing methods are used. Unfortunately, a similar property does not hold for the PH model (see Hougaard[42] for more discussion to this point).

3. Penalized AFT model

It seems that, in contrast to the PH model, it is much more difficult to develop methods for the AFT model with censored data which leave the baseline distribution completely unspecified, at the same time allow for inclusion of a high number of covariates and offer tools to compute predictive survival or hazard curves. On the other hand, it is obvious that parametric methods do not offer enough flexibility to model correctly survival data. Penalized smoothing forms a trade-off between the above two approaches. Indeed, a density or hazard of the baseline distribution using a large basis of parametric functions profits from the advantages of parametric methods while leaving the baseline distribution in practice largely unspecified. Among other things, this implies that prediction is easily accomplished.

The penalized AFT model of Komárek et al.[2] was motivated by the P-spline smoothing approach of Eilers and Marx[43] applied to the error density $f(e)$ in model (3). However, for reasons given in Komárek et al.[2] it was decided to replace P-splines by normal densities which results in a model (3) with

$$f(e) = \sum_{j=1}^{k} c_j \, \varphi_{\mu_j, \sigma_0^2}(e), \tag{5}$$

where $\varphi_{\mu_j,\sigma_0^2}(\cdot)$ is the normal density with mean μ_j and variance σ_0^2. The number of components k in expression (5) is fixed, however high (say around 30) to offer enough flexibility in the resulting density $f(e)$. The means $\mu_1, \ldots, \mu_k$ are fixed and play the role of knots in P-spline smoothing. These knots are assumed equidistant and to constitute a fine grid. The last fixed parameter of the model is the so called *basis standard deviation* σ_0. Further technical details can be found in Komárek et al.[2] Model (3) with (5) will be called *the mean penalized AFT model*.

The regression parameters $\boldsymbol{\beta}$, the intercept α, the scale σ and the weights $c_j, j = 1, \ldots, k$ are estimated by maximizing the penalized log-likelihood equal to the log-likelihood plus a penalty term to control the smoothness of the estimate of the error density $f(e)$. Asymptotic properties of the resulting estimates that serve as a basis for the inference are given in Komárek et al.[2] Finally, different models can be compared by means of Akaike's information criterion (AIC, Akaike[44]) with higher values indicating a better fitted model.

In most regression models, it is conventionally assumed that the covariates influence the mean, but it is presumed that it will *not* influence the scale parameter. With hindsight, this is simply one model choice and in many cases it may be untenable. Recently, there is interest in joint mean-covariance models in the context of longitudinal studies (Pourahmadi[45]; Pan and MacKenzie[46]). The original penalized AFT model of Komárek et al.[2] can be generalized in the same direction yielding *the mean-scale penalized AFT model*. With this generalization, we allow the scale parameter σ to depend on a vector of covariates, say $\boldsymbol{z}$, as

$$\sigma \equiv \sigma(\boldsymbol{z}) = \exp(\gamma_0 + \boldsymbol{\gamma}'\boldsymbol{z}), \tag{6}$$

For estimation of parameters γ_0 and $\boldsymbol{\gamma}$, the method of penalized maximum-likelihood used for the original mean penalized AFT model had to be adjusted only slightly.

Finally, we point out that the software to fit both mean and mean-scale penalized AFT models is available as a contributed package `smoothSurv` of the statistical package R[12] (free clone of S-PLUS) and is downloadable from its web page `http://www.R-project.org`.

4. Analysis of the Signal Tandmobiel® data

The penalized AFT models can be used to tackle the questions posed in Section 1 for the Signal Tandmobiel® data. For the sake of illustration, only teeth 14 and 15 will be considered. Further, we will tackle the question for these teeth separately, i.e. ignoring their correlation. In the discussion we indicate how the correlation between teeth can be incorporated in the analysis.

As response, for a particular child, we consider the age of emergence of a particular permanent tooth (14 or 15), recorded in years. For a better fit, we shifted the time origin of the AFT model to 5 years of age, i.e. by replacing T by $T - 5$ in the AFT model specification (3). Four mean penalized AFT models and two mean-scale penalized AFT models described in Table 1 were fitted. AIC's for these models are given in Table 2.

Firstly, the model with the interaction term $gender * dmf$ seems to fit the data best and the interaction term cannot be omitted. Secondly, the models where the scale parameter σ depends on covariates give a better fit, but for this only dmf seems to be necessary. These findings lead us to conclude that the model that describes satisfactory well the data while being kept as simple as possible is the model $gender * dmf/\text{scale}(dmf)$. The estimates for this model are given in Table 3. It is seen that $dmf = 1$ accelerates the emergence for both genders and also increases the variability of the emergence distribution.

4.1. **Predictive curves.**

The penalized AFT model has actually a parametric nature given the weights $c_1, \ldots, c_k$ in (5) are known. This makes it easy to compute predictive emergence curves or predictive hazards or densities for a given combination of covariates, say $\boldsymbol{x}^*$ and

$\boldsymbol{z}^*$. For instance, the predictive emergence curve is given by the relationship

$$\hat{F}(t \mid \boldsymbol{x}^*, \boldsymbol{z}^*) = \sum_{j=1}^{k} \hat{c}_j \Phi_{\mu_j, \sigma_0^2}\left(\frac{\log(t) - \hat{\alpha} - \hat{\boldsymbol{\beta}}' \boldsymbol{x}^*}{\hat{\sigma}(\boldsymbol{z}^*)}\right), \tag{7}$$

where $\Phi_{\mu_j, \sigma_0^2}(\cdot)$ denotes a cumulative distribution function of $N(\mu_j, \sigma_0^2)$. Predictive hazards or densities can be derived from (7).

For our data, predictive emergence curves based on the model $gender * dmf/\text{scale}(dmf)$ are shown in Figure 2 and predictive hazards in Figure 3. Further, Figure 2 shows also the non-parameric estimates of Turnbull[9] computed separately for each combination of covariates. It is seen that model-based emergence curves agree with the non-parametric estimates indicating the goodness-of-fit of our model. Further, the figures show that the difference between children with $dmf = 0$ and $dmf = 1$ is higher for boys than for girls and that the emergence process for boys is indeed postponed compared to girls.

Non-decreasing predictive hazard curves reflect the nature of the problem at hand. Indeed, it can be expected that, provided the tooth of a child has not emerged yet, the probability that the tooth will emerge increases with age.

4.2. **Comparison of emergence distributions between different groups.** While the model $gender * dmf/\text{scale}(dmf)$ gives a parsimonious description of emergence distributions for different groups of children and serves as a solid basis for prediction as was shown in the previous section, it is not suitable to provide simple p-values for a comparison of emergence distributions between e.g. boys and girls. Due to the fact that an interaction term $gender * dmf$ appeared to be significantly important, we could only provide a p-value for a multiple comparison of the four groups (girls with $dmf = 1$ and 0 and boys with $dmf = 1$ and 0).

To simply compare two distributions, while averaging the effect of other covariates, the mean AFT model with a univariate covariate x (i.e. either the model *gender* or the model

dmf) can be used together with a significance test for the group parameter. Additionally, it is possible to perform a test that compares two groups while controling for additional confounding variables (e.g. comparison of boys and girls while controling for dmf or vice versa). To do that, we perform significance tests of $\boldsymbol{\beta}$ parameters in the model $gender+dmf$.

The estimates of regression parameters $\boldsymbol{\beta}$ together with their standard errors in mentioned models are given in Table 4. The Wald tests of significance for each β parameter all yield p-values lower than 0.0001 which confirm the findings obtained previously that there is indeed a significant difference in emergence distributions of studied teeth between boys and girls and also between the group of children with $dmf = 0$ and $dmf = 1$. The difference remains both marginally and while controling for the other covariate.

The issue of the robustness of the AFT model against the omitted covariates is further illustrated in Table 4. The effect of *gender* remains almost unchanged in both models, *gender* and $gender + dmf$, and an analogous conclusion holds also for the effect of dmf.

4.3. **Conclusions.** It has been shown that the emergence processes of teeth 14 and 15 are significantly different between boys and girls and that the caries experience status of a primary predecessor, expressed by the dmf score, has a significant effect on the timing of emergence of permanent successors.

Predictive emergence curves have been drawn that can be used for diagnosis and treatment planning in paediatric dentistry. Further, it was found that the acceleration effect of caries experience on a primary predecessor on the timing of emergence of its successor was stronger for boys than for girls.

Finally, R scripts used to perform the analysis using the library `smoothSurv` can be found in the downloadable distribution of this library.

5. Discussion

Interval-censored data are not rare in medicine and certainly in dentistry, yet they are often not analysed with the proper techniques. Our article tried to give an overview of available approaches in a hope that their use will increase. Unfortunately, we have to admit that the daily use of relevant techniques is still complicated by insufficient ready-to-use software support.

Further, we should mention that Bayesian methods convey many attractive properties useful in the analysis of interval-censored data. Due to the fact that all unknown quantities are treated in the same manner by Bayesian techniques one can include (unknown since censored) true event-times in the model as additional parameters while focusing the inference only on parameters of interest (e.g. regression parameters). Necessarily, there is no additional complexity when analysing interval-censored data compared to right-censored data. However, a review of Bayesian techniques for survival analysis would go beyond the scope of this text.

Furthermore, we have shown how the accelerated failure time model with a flexible error distribution estimated using the penalized maximum-likelihood, could be used for the analysis of interval-censored data.

Finally, we admit that our analysis of the Signal Tandmobiel® data did not fully take into account the complexity of the problem. Indeed, we ignored the correlation between tooth 14 and tooth 15. It is somewhat typical that dental problems immediately become quite complex due to the fact that teeth are associated since they are spatially close to each other. To resolve this problem one could treat the two teeth jointly by using an AFT model with a bivariate error distribution (see e.g. Pan and Connett[40]). Another possibility is to extend the AFT model with a random intercept for mouth thereby introducing a frailty component to the model (see Pan and Louis[39] for possible semi-parametric approach). For this purpose,

a parametric AFT model with the random intercept following a normal distribution and the baseline survival time following either a log-normal, a log-logistic or a Weibull distribution can be fitted in R using the function `survreg` and in S-PLUS using the function `SurvReg`.

Acknowledgements

We want to thank Prof. N. Keiding for his suggestion to model also the scale parameter in the penalized AFT model as a function of covariates.

The research presented here was financially supported by Research Grant OE/03/29, Katholieke Universiteit Leuven. The authors further acknowledge support from the Interuniversity Attraction Poles Program P5/24 – Belgian State – Federal Office for Scientific, Technical and Cultural Affairs.

The Signal Tandmobiel® project comprises the following partners: D. Declerck (Dental School, Catholic University Leuven), L. Martens (Dental School, University Ghent), J. Vanobbergen (Oral Health Promotion and Prevention, Flemish Dental Association), P. Bottenberg (Dental School, University Brussels), E. Lesaffre (Biostatistical Centre, Catholic University Leuven), K. Hoppenbrouwers (Youth Health Department, Catholic University Leuven; Flemish Association for Youth Health Care). Data collection was supported by Unilever, Belgium.

References

[1] Vanobbergen J, Martens L, Lesaffre E, and Declerck D. The Signal-Tandmobiel® project – a longitudinal intervention health promotion study in Flanders (Belgium): baseline and first year results. *European Journal of Paediatric Dentistry* 2000; **2**: 87–96.

[2] Komárek A, Lesaffre E, and Hilton JF. Accelerated failure time model for arbitrarily censored data with smoothed error distribution. *IAP Technical Report 0451* 2004; `URL http://www.stat.ucl.ac.be/IAP.`

[3] Rücker G, and Messerer D. Remission duration: an example of interval-censored observations. *Statistics in Medicine* 1988; **7**: 1139–1145.

[4] Law CG, and Brookmeyer R. Effects of mid-point imputation on the analysis of doubly censored data. *Statistics in Medicine* 1992; **11**: 1569–1578.

[5] Odell PM, Anderson KM, and D'Agostino RB. Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. *Biometrics* 1992; **48**: 951–959.

[6] Dorey FJ, Little RJ, and Schenker N. Multiple imputation for threshold-crossing data with interval censoring. *Statistics in Medicine* 1993; **12:** 1589–1603.

[7] Kaplan EL, and Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 1958; **53**: 457–481.

[8] Peto R. Experimental survival curves for interval-censored data. *Applied Statistics* 1973; **22**: 86–91.

[9] Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society, Series B* 1976; **38**: 290–295.

[10] Dempster AP, Laird NM, and Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* 1977; **39**: 1–38.

[11] Kooperberg C, and Stone CJ. Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics* 1992; **1**: 301–328.

[12] R Development Core Team *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing, 2004; ISBN 3-900051-00-3, `URL http://www.R-project.org.`

[13] Rosenberg PS. Hazard function estimation using B-splines. *Biometrics* 1995; **51**: 874–887.

[14] Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother. Rep.* 1966; **50**: 163–170.

[15] Gehan EA. A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* 1965; **52**: 203–223.

[16] Peto R, and Peto J. Asymptotically efficient rank-invariant test procedures (with Discussion). *Journal of the Royal Statistical Society, Series A* 1972; **135**: 185–206.

[17] Prentice RL. Linear rank tests with right censored data. *Biometrika* 1978; **65**: 167–179.

[18] Pepe MS, and Fleming TR. Weighted Kaplan-Meier statistics: a class of distance tests for censored survival data. *Biometrics* 1989; **45**: 497–507.

[19] Mantel N. Ranking procedures for arbitrarily restricted observations. *Biometrics* 1967; **23**: 65–78.

[20] Self SG, and Grossman EA. Linear rank tests for interval-censored data with application to PCB levels in adipose tissue of transformer repair workers. *Biometrics* 1986; **42**: 521–530.

[21] Finkelstein DM. A proportional hazards model for interval-censored failure time data. *Biometrics* 1986; **42**: 845–854.

[22] Petroni GR, and Wolfe RA. A two-sample test for stochastic ordering with interval-censored data. *Biometrics* 1994; **50**: 77–87.

[23] Fay MP. Rank invariant tests for interval censored data under grouped continuous model. *Biometrics* 1996; **52**: 811–822.

[24] Fang H-B, Sun J, and Lee M-LT. Nonparametric survival comparisons for interval-censored continuous data. *Statistica Sinica* 2002; **12**: 1073–1083.

[25] Pan W. A two-sample test with interval censored data via multiple imputation. *Statistics in Medicine* 2000; **19**: 1–11.

[26] Cox DR. Regression models and life-tables (with Discussion). *Journal of the Royal Statistical Society, Series B* 1972; **34**: 187–220.

[27] Cox DR. Partial likelihood. *Biometrika* 1975; **62**: 269–276.

[28] Breslow NE. Covariance analysis of censored survival data. *Biometrics* 1974; **30**: 89–99.

[29] Satten GA. Rank-based inference in the proportional hazards model for interval censored data. *Biometrika* 1996; **83**: 355–370.

[30] Satten GA, Datta S, and Williamson JM. Inference based on imputed failure times for the proportional hazards model with interval-censored data. *Journal of the American Statistical Association* 1998; **93**: 318–327.

[31] Goggins WB, Finkelstein DM, Schoenfeld DA, and Zaslavsky AM. A Markov chain Monte Carlo EM algorithm for analyzing interval-censored data under the Cox proportional hazards model. *Biometrics* 1998; **54**: 1498–1507.

[32] Pan W. A multiple imputation approach to Cox regression with interval-censored data. *Biometrics* 2000; **56**: 199–203.

[33] Goetghebeur E, and Ryan L. Semiparametric regression analysis of interval-censored data. *Biometrics* 2000; **56**: 1139–1144.

[34] Kooperberg C, and Clarkson DB. Hazard regression with interval-censored data. *Biometrics* 1997; **53**: 1485–1494.

[35] Betensky RA, Lindsey JC, Ryan LM, and Wand MP. Local EM estimation of the hazard function for interval-censored data. *Biometrics* 1999; **55**: 238–245.

[36] Kalbfleisch JD, and Prentice RL. *The Statistical Analysis of Failure Time Data, 2nd Edition.* Hoboken: John Wiley & Sons, 2002.

[37] Rabinowitz D, Tsiatis A, and Aragon J. Regression with interval-censored data. *Biometrika* 1995; **82**: 501–513.

[38] Betensky RA, Rabinowitz D, and Tsiatis AA. Computationally simple accelerated failure time regression for interval censored data. *Biometrika* 2001; **88**: 703–711.

[39] Pan W, and Louis TA. A linear mixed-effects model for multivariate censored data. *Biometrics* 2000; **56**: 160–166.

[40] Pan W, and Connett JE. A multiple imputation approach to linear regression with clustered censored data. *Lifetime Data Analysis* 2001; **7**: 111–123.

[41] Pan W, and Kooperberg C. Linear regression for bivariate censored data via multiple imputation. *Statistics in Medicine* 1999; **18**: 3111–3121.

[42] Hougaard P. Fundamentals of survival data. *Biometrics* 1999; **55**: 13–22.

[43] Eilers PHC, and Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science* 1996; **11**: 89–121.

[44] Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974; **AC-19**: 716–723.

[45] Pourahmadi M. Joint mean-covariance models with applications to longitudinal data: Unconstrained parametrisation. *Biometrika* 1999; **86**: 677–690.

[46] Pan J, and MacKenzie G. On modelling mean-covariance structures in longitudinal studies. *Biometrika* 2003; **90**: 239–244.

TABLE 1. Description of fitted models.

Model	Covariate $\boldsymbol{x}$	Covariate $\boldsymbol{z}$
$gender$	$(gender)$	—
dmf	(dmf)	—
$gender + dmf$	$(gender,\ dmf)'$	—
$gender * dmf$	$(gender,\ dmf,\ gender \times dmf)'$	—
$gender * dmf/\text{scale}(dmf)$	$(gender,\ dmf,\ gender \times dmf)'$	(dmf)
$gender * dmf/\text{scale}(gender * dmf)$	$(gender,\ dmf,\ gender \times dmf)'$	$(gender,\ dmf,\ gender \times dmf)'$

TABLE 2. Akaike's information criteria for different models.

Model	Tooth 14	Tooth 15
$gender$	-5532.59	-4551.57
dmf	-5538.03	-4549.93
$gender + dmf$	-5494.51	-4526.85
$gender * dmf$	-5491.47	-4522.76
$gender * dmf/\text{scale}(dmf)$	-5468.61	-4506.66
$gender * dmf/\text{scale}(gender * dmf)$	-5467.67	-4507.59

TABLE 3. Estimates (standard errors) for the model $gender * dmf/\text{scale}(dmf)$.

Parameter	Tooth 14	Tooth 15
α	1.7734 (0.0073)	1.9143 (0.0091)
$\beta(gender)$	-0.0931 (0.0099)	-0.0803 (0.0110)
$\beta(dmf)$	-0.0990 (0.0116)	-0.0773 (0.0125)
$\beta(gender * dmf)$	0.0401 (0.0166)	0.0473 (0.0172)
γ_0	-1.5613 (0.0219)	-1.6121 (0.0351)
$\gamma(dmf)$	0.2144 (0.0307)	0.2415 (0.0399)

TABLE 4. Estimates (standard errors) for models *gender*, dmf and *gender*+ dmf.

Parameter	Model *gender* or dmf	Model *gender* + dmf
	Tooth 14	
$\beta(gender)$	-0.0740 (0.0080)	-0.0766 (0.0081)
$\beta(dmf)$	-0.0729 (0.0086)	-0.0741 (0.0085)
	Tooth 15	
$\beta(gender)$	-0.0564 (0.0085)	-0.0594 (0.0087)
$\beta(dmf)$	-0.0613 (0.0089)	-0.0628 (0.0090)

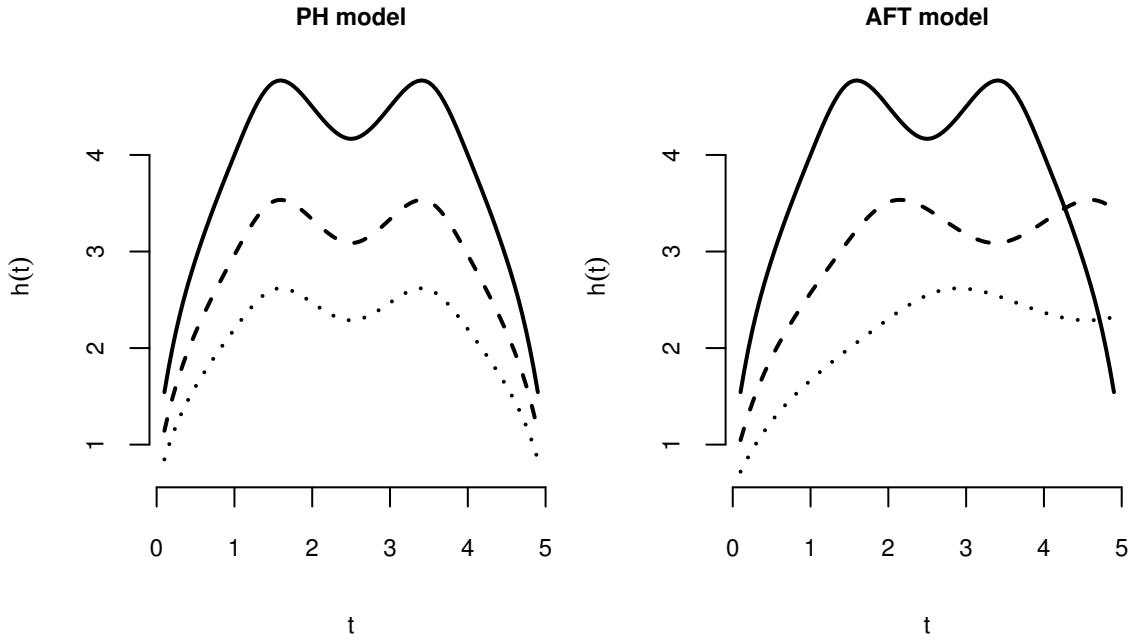

FIGURE 1. Effect of PH and AFT assumption on a hypothetical baseline hazard function (solid line) for a univariate covariate x taking a value of 0.6 (dashed line) and 1.2 (dotted line) with regression parameter $\beta = -0.5$ for the PH model and $\beta = 0.5$ for the AFT model.

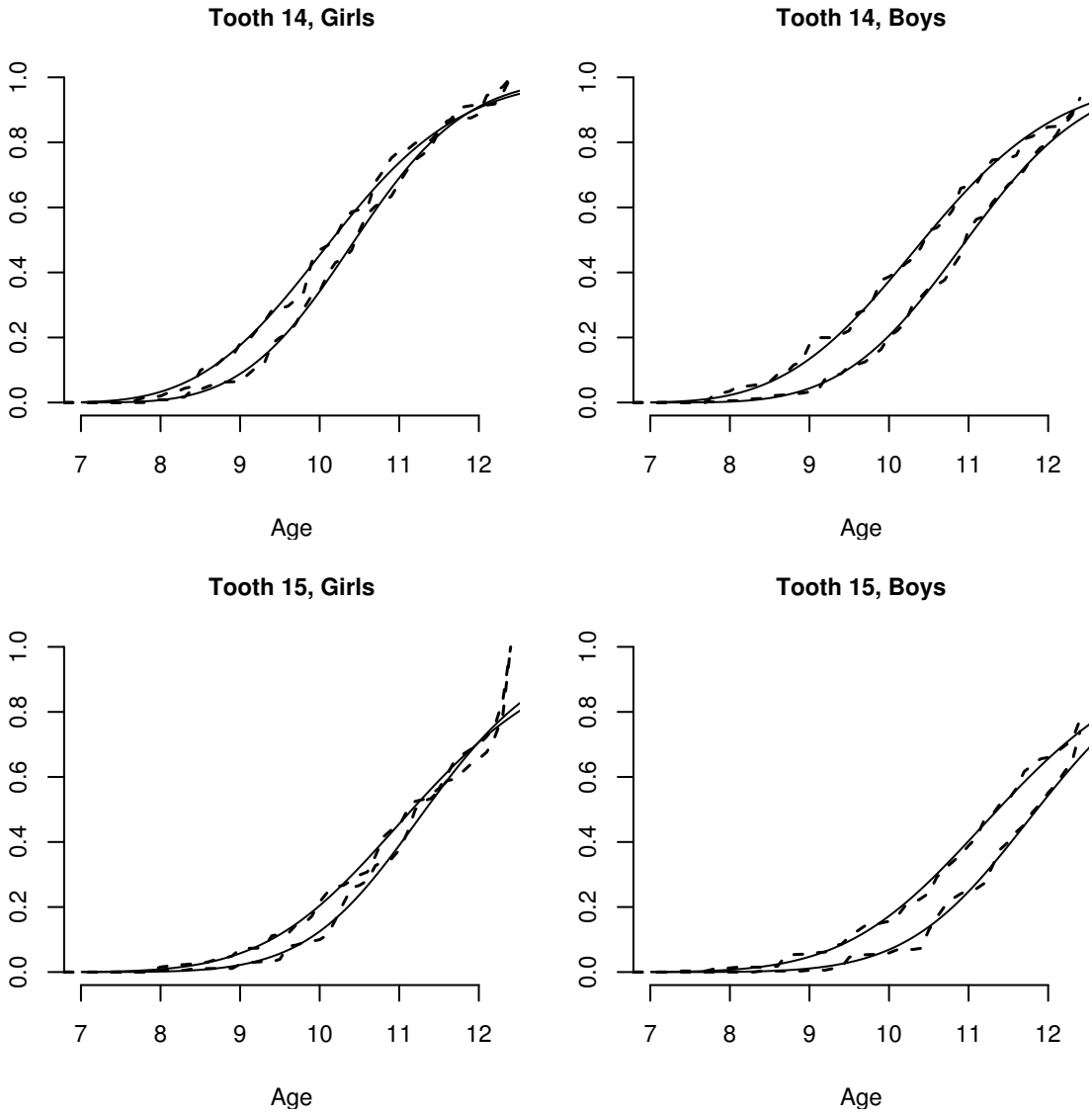

FIGURE 2. Predictive emergence curves: solid lines for curves based on the model $gender * dmf/\text{scale}(dmf)$ (on each plot: left curve for $dmf = 1$, right curve for $dmf = 0$), dashed line for a non-parametric estimate of Turnbull.

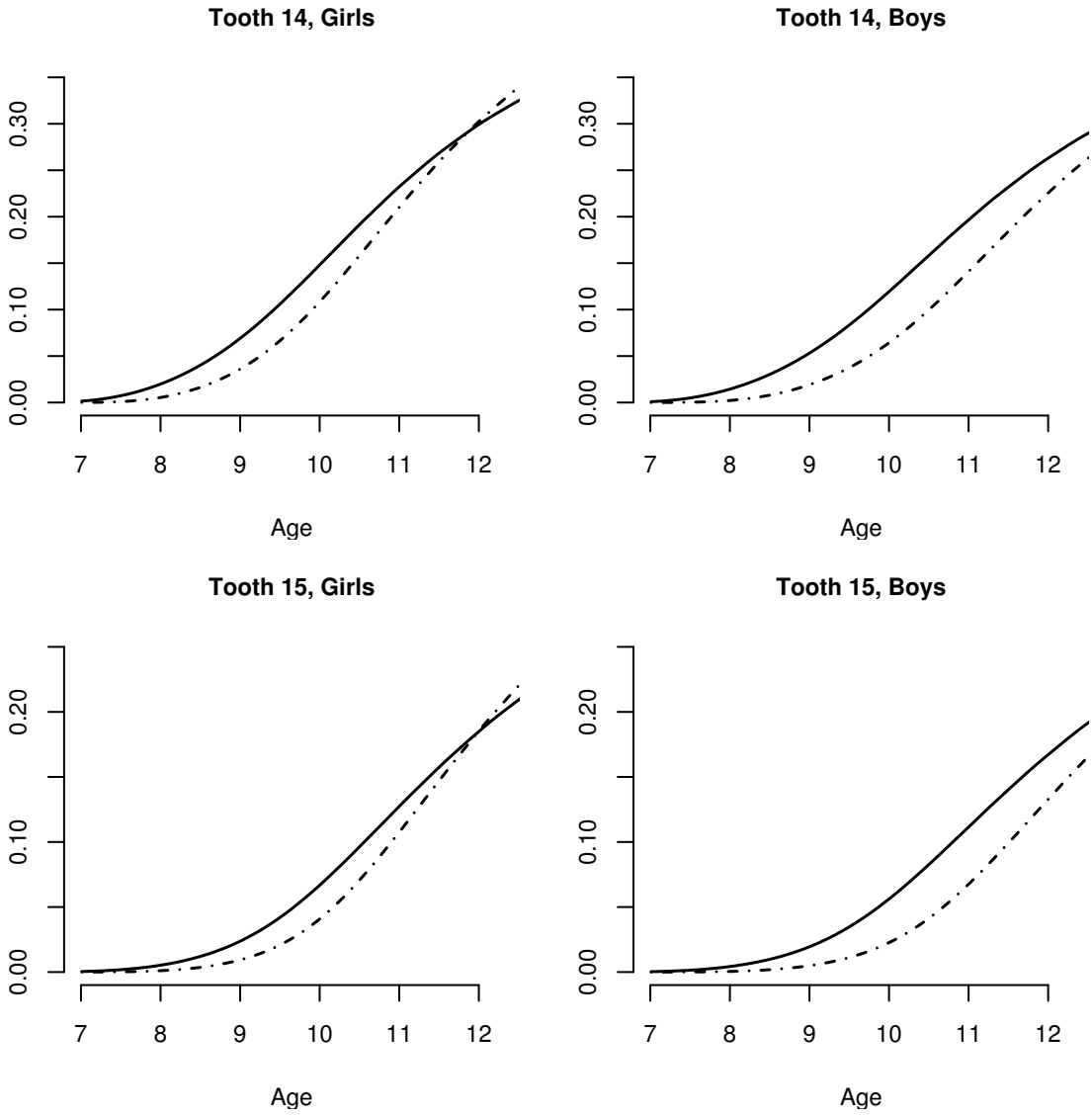

FIGURE 3. Predictive hazard curves of emergence based on the model $gender * dmf/\text{scale}(dmf)$: solid line for $dmf = 1$, dotted-dashed line for $dmf = 0$.