Received: 10 May 2022 | Revised: 25 August 2022 | Accepted: 13 January 2023

DOI: 10.1111/stan.12287

ORIGINAL ARTICLE

WILEY

Inference in the presence of likelihood monotonicity for proportional hazards regression

John E. Kolassa[1] | **Juan Zhang**[2]

[1]Department of Statistics, Rutgers University, New Brunswick, New Jersey, USA

[2]Statistical Sciences, AbbVie, Inc., North Chicago, Illinois, USA

Correspondence
John E. Kolassa, Department of Statistics, Rutgers University, New Brunswick, NJ, USA.
Email: kolassa@stat.rutgers.edu

Funding information
United States National Science Foundation

Proportional hazards are often used to model event time data subject to censoring. Samples involving discrete covariates with strong effects can lead to infinite maximum partial likelihood estimates. A methodology is presented for eliminating nuisance parameters estimated at infinity using approximate conditional inference. Of primary interest is testing in cases in which the parameter of primary interest has a finite estimate, but in which other parameters are estimated at infinity.

KEYWORDS
conditional inference, likelihood monotonicity, proportional hazards regression

1 | INTRODUCTION

The proportional hazards regression model (Cox, 1972) is commonly used to model the dependence between time to an event and various covariates. When times to event have continuous distributions, this model constrains the hazard (defined as the negative of the derivative of the log of the survival function) to be a baseline function times the exponential of a linear combination of the covariates, and generally the practitioner wants to learn about the coefficients in this linear relationship, while imposing minimal conditions on the baseline hazard function. Often the event times for some experimental subjects are not observed precisely, but are observed to exceed some time, called the censoring time; this phenomenon is known as right censoring.

In order to eliminate the effect of the unknown baseline hazard function from the analysis, inference is often performed by constructing a function from the full likelihood, formed by discarding terms containing information on the gaps between succeeding failures (Cox, 1972). This resulting function, called the partial likelihood, has many properties in common with the full likelihood, and is often used in the same way that the full likelihood is used; for example, parameters are often estimated by maximizing the partial likelihood, standard errors are often calculated from the second derivative of the log of the partial likelihood, and the change in the maximized likelihood as one moves from a larger model to a smaller model nested within it is often used for testing.

This partial likelihood is exactly the same as the likelihood that arises from a multinomial regression model with a dataset derived from the proportional hazards dataset, to be described below, and, in turn, the multinomial distribution (and hence its likelihood) will be represented exactly by a certain conditional logistic regression model.

Datasets with a monotone relationship between a covariate and event times give rise to likelihood functions that do not have a maximum; instead, there exists one or more linear combinations of the parameters such that as such a linear combination is increased to infinity, the likelihood continues to increase. This monotonicity in the likelihood complicates estimation and testing of regression parameters. Logistic regression, multinomial regression, and proportional hazards regression models share this difficulty of potential likelihood monotonicity. During the iterative process of fitting the model, an important numerical calculation (generally first the inversion of the second derivative matrix of the log likelihood) becomes impossible, and the algorithm stops. The most naive response to the problem of likelihood monotonicity is to report numerical results at the last computable step, and to exit with a warning message. In this context, one might run the standard algorithm until it fails numerically. An advantage of this approach is that it produces confidence intervals and *p*-values generally, although not always, close to our approach suggested here. Unfortunately, performance and stability not guaranteed.

A more sophisticated approach involves regularization, with a penalty function in frequentist inference. Alternatively, one might perform a regularized approach (Heinze & Schemper, 2001), by multiplying the partial likelihood by the Jeffreys prior (Firth, 1993), by a matching prior (Zhang & Kolassa, 2013), or using a proper Bayesian prior. Other regularization functions yield estimates with different characteristics (Pagui & Colosimo, 2020). An advantage of regularization is numerical stability. A finite maximizer for the penalized likelihood always exists (Heinze & Schemper, 2001); however, numerical attempts to find this maximizer, as implemented in SAS/STAT (SAS Institute, Inc., 2019) and in R (Heinze & Ploner, 2018), sometimes fail. Furthermore, one might use the penalized likelihood as one might use an unpenalized likelihood (Heinze & Schemper, 2001), to create Wald and likelihood ratio tests; critical values and *p*-values are calculated using the Gaussian distribution as usual. This Gaussian approximation may be justified theoretically (Hjort, 1986).

Linear optimization may be used to diagnose and adjust for this monotonicity in conditional logistic regression models (Kolassa, 1997), and these techniques may be applied to multinomial regression models (Kolassa, 2016). This manuscript extends these linear optimization methods to proportional hazards regression to provide approximate inference. Of primary interest is testing in cases in which the parameter of primary interest has a finite estimate, but in which other parameters are estimated at infinity. Concerns about testing include considerations of whether nominal test level matches actual test level, and of power. Effects on the quality of estimates for parameters are presented for comparison with the regularization method (Heinze & Schemper, 2001).

Advantages of this proposed method are that resulting confidence intervals for parameters estimated at infinity correctly have one infinite endpoint; in contrast, competing methods using a likelihood penalty(Heinze & Schemper, 2001) give intervals with finite endpoints, and these intervals exclude parameter values indicated by the partial likelihood as being more likely than some of those included. This finiteness of confidence intervals for parameters estimated at infinity is entirely appropriate for an explicitly Bayesian solution, but is deeply problematic for an analysis that is not intended to rely heavily on a subjective prior. Furthermore, the proposed method more closely tracks the paradigm of conditional inference.

In the remainder of this paper, Section 2 presents a motivating example and considers the impact of extreme data sets on logistic regression, Section 3 discusses models more complicated than logistic regression, Section 4 develops diagnostic tools for determining the presence of infinite estimates in proportional hazards regression, Section 5 presents our proposed technique for avoiding infinite parameter estimates, Section 6 presents some examples and numerical assessments of our method, and Section 7 presents some conclusions.

2 | THE MOTIVATING EXAMPLE

Consider a study of 103 subjects with invasive ductal carcinomas (Lösch et al., 1998). A subset of this data, featuring the time (in months) to 26 deaths or 74 censorings, and covariates tumor stage (T), nodal status (N), histological grading (G), and cathepsin D immunoreactivity (CD), is analyzed below (Heinze & Schemper, 2001). We consider assessment of the effect of tumor stage T on survival.

The package `survival` (Therneau, 2015) in the program R (R Core Team, 2020) gives the parameter estimates in Table 1. The main effect of G is quite large. Furthermore, the software provides a warning message indicating that convergence fails. This is generally (but not always) because the true maximizer is at infinity. Problematically, the proportional hazards algorithm fails to converge, because the partial likelihood is monotone in the coefficient of G. Such monotonicity is more likely in a dataset with more covariates, and a dataset with dichotomous covariates. Removing G gives the results in Table 2, but the T estimate changes quite a bit, implying that G is a strong predictor. Hence dropping G may lead to bias.

2.1 | Parallels with logistic regression

Analyzing this dataset by modeling the probability of death using logistic regression has the same problem. Consider random variables Y_j taking the value 1 if the subject is observed to die, and taking the value 0 otherwise. Model these random variables as independent, with

TABLE 1 Results of proportional hazards regression for breast cancer survival

Variable	Estimate	SE	p
T	1.279	0.502	.0109
N	0.946	0.425	.0260
G	18.421	4,316.768	.9966
CD	0.400	0.444	.3670

TABLE 2 Results of proportional hazards regression for breast cancer survival, without problematic covariate

Variable	Estimate	SE	*p*
T	1.560	0.501	.002
N	1.134	0.432	.009
CD	0.521	0.450	.247

TABLE 3 Results of logistic regression for predicting event before 22 months

Variable	Estimate	SE	*p*
T	1.6105	0.8633	.0621
N	0.8583	0.7101	.2268
G	16.9082	20,40.4185	.9934
CD	0.5595	0.7415	.4506
(Intercept)	−20.2393	2,040.4184	.9921

$$P\left[Y_j = 1\right] = \exp(\boldsymbol{z}_j\boldsymbol{\gamma})/(1+\exp(\boldsymbol{z}_j\boldsymbol{\gamma})), \quad P\left[Y_j = 0\right] = 1/(1+\exp(\boldsymbol{z}_j\boldsymbol{\gamma})), \tag{1}$$

for $\boldsymbol{x}_j$ the vectors containing covariates as given in Table 1, and $\boldsymbol{z}_j = (\boldsymbol{x}_j, 1)$. (Note that the intercept term is added as the last, rather than first, parameter, in order to keep the position of other covariates constant). In this case, inference on the parameter for T, γ_1, is desired, without specifying values for the other parameters. Sufficient statistics for the parameter $\boldsymbol{\gamma}$ are given by

$$\boldsymbol{V} = \boldsymbol{Z}^\top \boldsymbol{Y},$$

for $\boldsymbol{Z}$ the matrix with rows $\boldsymbol{z}_j$. Model (1) forms a natural exponential family, and so the distribution of $V_1|V_2, \ldots, V_5$ depends on the parameter of interest γ_1, and not on any of the other model parameters $\gamma_2, \ldots, \gamma_5$. Computations to calculate this conditional distribution and complete estimation and testing can be lengthy, and more often, the estimate $\hat{\boldsymbol{\gamma}}$ maximizing the full likelihood is presented. For moderately large samples, the distribution of $\hat{\boldsymbol{\gamma}}$ is approximately multivariate normal. The approximating distribution has expectation $\boldsymbol{\gamma}$, and a variance matrix calculated from $\ell''(\boldsymbol{\gamma})$, with ℓ the logistic regression log likelihood; since $\boldsymbol{\gamma}$ is unknown, the estimate $\hat{\boldsymbol{\gamma}}$ is substituted into the variance formula.

Logistic regression gives the results in Table 3. As noted above, the logistic regression results contain an additional parameter, labeled Intercept. Large parameter estimates appear in Table 1 in the same places as in Table 3. In this case, no finite maximum likelihood estimators exist, and substituting directly into the variance formula is impossible.

Inferential issues arise in this logistic regression, because the estimate of the parameter associated with *G* is infinite. If all parameter estimates were finite, standard normal approximations to *p*-values and confidence intervals could be employed. Were the only parameter estimated at infinity the interest parameter, *p*-value calculations employing a continuity correction may be performed using standard large-sample approximate multivariate normal methods. That is, the conditional *p*-value for testing the null hypothesis $\gamma_1 = \gamma_1^\circ$ is

$$2\min(\underset{\gamma_1^\circ}{P}\left[V_1 \geq v_1^- | V_2, \ldots, V_5\right], \underset{\gamma_1^\circ}{P}\left[V_1 \leq v_1^+ | V_2, \ldots, V_5\right]),$$

where v_1^+ and v_1^- represent the observed values for V_1, moved higher and lower by half the minimal spacing of values in the conditional sample space of $V_1|V_2, \ldots, V_5$. One of the two probabilities is 1, and for the other, the value v_1^+ or v_1^- of V_1 at which the maximum likelihood estimator must be calculated sits strictly between its largest and smallest possible values, and so a finite maximizer exists.

Furthermore, the relationship between null hypothesis γ_1° and the p-value can be inverted to give a confidence interval. If γ_1 is involved a contrast estimated at infinity, then the observed value of V_1 is at the end of the conditional sample space for $V_1|V_2, \ldots, V_5$, and hence the confidence interval has one end point $\pm\infty$, with the other end point determined by solving the equation

$$\underset{\gamma_1^\circ}{P}\left[V_1 \geq v_1^- | V_2, \ldots, V_5\right] = \alpha/2. \tag{2}$$

$$\underset{\gamma_1^\circ}{P}\left[V_1 \leq v_1^+ | V_2, \ldots, V_5\right] = \alpha/2. \tag{3}$$

The appropriate value v_1^+ or v_1^- is associated with a maximum likelihood estimator $\hat{\boldsymbol{\gamma}}$ that is not infinite along any contrast involving γ_1.

Proportional hazards regression models share this same phenomenon of infinite parameter estimates leading to inference problems when the parameters not of interest is estimated at infinity. Hence this paper only considers cases in which infinite parameter estimates do not involve the parameter of interest.

Inference in such logistic regression cases may be made by considering the conditional distribution $V_1|V_2, \ldots, V_5$, creating a smaller dataset with the same conditional distribution that avoids infinite estimates, and applying normal approximation methods, or higher-order inference to provide more accurate approximation methods, to the conditional distribution (Kolassa, 1997). This technique extends to multinomial regression (Kolassa, 2016). Later sections of this manuscript extend this technique to proportional hazards regression. This suggestion represents an extension from a canonical exponential family to a curved exponential family, and approximates the exact likelihood by a partial likelihood.

Logistic regression models with a large components of the parameter vector $\boldsymbol{\gamma}$ are likely to lead to datasets yielding infinite estimates. Binary covariates, and particularly those with a preponderance of one of the two possible values, also more often lead to infinite estimates.

Other authors have discussed this problem generally (Mansournia, Geroldinger, Greenland, & Heinze, 2017), and have presented a solution to the logistic regression problem by regularizing using a prior (Heinze & Ploner, 2003). Computational tools exist for a fully Bayesian solution (Wu, de Castro, Schifano, & Chen, 2018). A variety of exact and regularized asymptotic approaches have been surveyed (Heinze & Puhr, 2010).

3 | MODELS BEYOND LOGISTIC REGRESSION

This section reviews the multinomial, logistic, and proportional hazards regression models. It also reviews the solution to the problem of monotone likelihood in multinomial regression using conditional logistic regression (Kolassa, 2016). This section further describes how the partial likelihood from the proportional hazards model may be expressed as the likelihood from a multinomial model.

3.1 | Multinomial regression

Suppose that $\mathcal{P}$ multinomial trials are observed; for trial $m \in \{1, \ldots, \mathcal{P}\}$, a choice D_m from the set $\mathcal{A}_m$ of alternatives is observed, with probability

$$P\left[D_m = d_m\right] = \exp(\boldsymbol{x}_{md_m}\boldsymbol{\beta}) / \sum_{k \in \mathcal{A}_m} \exp(\boldsymbol{x}_{mk}\boldsymbol{\beta}).$$

Here $\boldsymbol{x}_{mj} \in \Re^{\mathcal{D}}$ are covariate vectors associated with each of the alternatives. The likelihood is given by

$$P_{\beta}\left[Y_{mj} = y_{mj} \forall m, j\right] = L(\boldsymbol{\beta}) = \prod_{m \in \mathcal{C}} \exp(\boldsymbol{x}_{mD_m}\boldsymbol{\beta}) / \sum_{k \in \mathcal{A}_m} \exp(\boldsymbol{x}_{mk}\boldsymbol{\beta}), \tag{4}$$

for

$$\mathcal{C} = \{1, \ldots, \mathcal{P}\}.$$

The log likelihood for this model is

$$\ell(\boldsymbol{\beta}) = \boldsymbol{U}\boldsymbol{\beta} - \sum_{m \in \mathcal{C}} \log\left(\sum_{j \in \mathcal{A}_m} \exp(\boldsymbol{x}_{ji}\boldsymbol{\beta})\right), \tag{5}$$

for $\boldsymbol{U} = \sum_{j=1}^{\mathcal{P}} \boldsymbol{x}_{jD_j}$, a sufficient statistic for $\boldsymbol{\beta}$.

Multinomial regression models yield infinite estimates estimates under similar circumstances for logistic regression. That is, infinite estimates become more likely for large number of unbalanced binary covariates, and with large values some components of the model parameter vector $\boldsymbol{\beta}$.

3.2 | The multinomial regression model as a special case of conditional logistic regression

This subsection presents the formulation of a multinomial regression model as a conditional logistic regression model. Let h_i be the number of entries in $\mathcal{A}_i$. The design matrix for the conditional logistic regression recasting the multinomial regression of Section 3.1 is

$$\begin{array}{c} \\ h_1 \text{ rows} \\ \\ \\ \\ h_2 \text{ rows} \\ \\ \\ \\ \vdots \\ h_{\mathcal{P}} \text{ rows} \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccc} \mathcal{D} \text{ columns} & & \multicolumn{3}{c}{\mathcal{P} \text{ columns}} \end{array} \\ \begin{pmatrix} \boldsymbol{x}_{11} & 1 & 0 & \ldots & 0 \\ \boldsymbol{x}_{12} & 1 & 0 & \ldots & 0 \\ \vdots & 1 & \vdots & \vdots & \vdots \\ \boldsymbol{x}_{1h_1} & 1 & 0 & \ldots & 0 \\ \boldsymbol{x}_{21} & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{x}_{2h_2} & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ \boldsymbol{x}_{\mathcal{P}1} & 0 & 0 & \ldots & 1 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ \boldsymbol{x}_{\mathcal{P}h_{\mathcal{P}}} & 0 & 0 & \ldots & 1 \end{pmatrix} \end{array}. \tag{6}$$

That is, the first $\mathcal{D}$ columns are constructed by stacking the h_1 covariate vectors for the first set of choices, followed by the h_2 covariate vectors associated with the second set of choices, and on, until the last $h_{\mathcal{P}}$ rows contain the last $h_{\mathcal{P}}$ covariate vectors. This forms a matrix with $h_1 + \cdots + h_{\mathcal{P}}$ rows and $\mathcal{D}$ columns. To this matrix is added additional $\mathcal{P}$ columns on the right. The first of these is a vector with all zeros, except for 1 in the first h_1 positions. The second is a vector of all zeros, except for 1 in positions $h_1 + 1$ through $h_1 + h_2$; that is, 1 starts just after 1 ended in the previous column. These repeat to fill the remainder of the last $\mathcal{P}$ columns.

Let $\boldsymbol{Z}$ denote the matrix in (6). Let $\boldsymbol{z}_j$ be row vector representing row j of $\boldsymbol{Z}$. Let $\boldsymbol{\tau}$ be an arbitrary row vector in $\Re^{\mathcal{P}}$, and $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \boldsymbol{\tau})$. Then, if $\boldsymbol{Y}$ is a random vector of independent responses, satisfying (1), and if

$$\begin{pmatrix} \boldsymbol{W} \\ \boldsymbol{U} \end{pmatrix} = \boldsymbol{Z}^{\top}\boldsymbol{Y},$$

then the distribution of $\boldsymbol{W}$ conditional on $\boldsymbol{U}$ is the same as the distribution of the sufficient statistic from a multinomial experiment.

Throughout this paper, inference for a proportional hazards model is related to conditional inference on an associated logistic regression model. The proportional hazards model parameter vector will be denoted $\boldsymbol{\beta}$. The associated logistic regression model has parameters corresponding to the same covariates, plus one or more intercept terms; this parameter vector is denoted $\boldsymbol{\gamma}$.

3.3 | Proportional hazards regression partial likelihood in relation to multinomial regression

The proportional hazards regression model considers a set of times $T_1, \ldots, T_{\mathcal{K}}$ at which $\mathcal{K}$ patients have an event. Assume that there exist vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\mathcal{K}}$ such that

$$\frac{d}{dt}\log\left(P\left[T_j \geq t\right]\right) / \frac{d}{dt}\log(P\left[T_k \geq t\right]) = \exp((\boldsymbol{x}_j - \boldsymbol{x}_k)\boldsymbol{\beta}) \text{ for all } j, k.$$

Suppose further that there exist censoring times C_k, either fixed, or independent of T_k, such that one may observe only $S_k = \min(T_k, C_k)$, and indicators δ_k, taking the value 1 if $T_k \leq C_k$, or 0 if $T_k > C_k$. One might estimate $\boldsymbol{\beta}$ by maximizing the partial likelihood (Cox, 1972), also given by

$$P_{\beta}\left[Y_{mj} = y_{mj} \forall m, j\right] = L(\boldsymbol{\beta}) = \prod_{m \in C} \pi_{mD_m}(\boldsymbol{\beta}),$$

for

$$\pi_{mj}(\boldsymbol{\beta}) = \frac{\exp(\boldsymbol{x}_{mj}\boldsymbol{\beta})}{\sum_{k \in \mathcal{A}_m} \exp(\boldsymbol{x}_{mk}\boldsymbol{\beta})},$$

as in (4). In this case, $C = \{k | \delta_k = 1\}$ (that is, this is the set of individuals having the event), $\mathcal{A}_m = \{a \in \{1, \ldots, \mathcal{K}\} | S_a \geq S_m\}$, the individuals at risk at time S_m, and $D_m = m$. The log partial likelihood is given by

$$\ell(\boldsymbol{\beta}) = \boldsymbol{U}\boldsymbol{\beta} - \sum_{m \in C} \log\left(\sum_{k \in \mathcal{A}_m} \exp(\boldsymbol{x}_{mk}\boldsymbol{\beta})\right), \tag{7}$$

for

$$\boldsymbol{U} = \sum_{j=1}^{\mathcal{P}} \boldsymbol{x}_{jD_j}. \tag{8}$$

Note the similarity between (7) and (5). The difference is that the second term in (7) is data dependent, in that the sets $\mathcal{A}_m$ depend on what happened in earlier steps.

Furthermore, similar characteristics making infinite estimates likely for the logistic and multinomial regression models also make infinite proportional hazards estimates more likely; these characteristics include a large number of binary with unbalanced binary covariates. Additionally, a large proportion of censored observations make infinite estimates more likely.

4 | INFINITE ESTIMATES

This section builds techniques for determining components of a proportional hazards parameter vector not having a finite maximum partial likelihood estimator. Extreme fitted probabilities may be determined by the following theorem (Kolassa, 1997):

Theorem 1. *Suppose that matrix $\boldsymbol{Z}$ is of full rank. Take $\boldsymbol{u} \in \Re^{\mathcal{K}}$ and $\boldsymbol{r}$ and $\boldsymbol{s}$ are column vectors of non-negative numbers such that*

$$\boldsymbol{u} = (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top (\boldsymbol{s} - \boldsymbol{r}), \tag{9}$$

$$(\boldsymbol{u}^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top - \boldsymbol{n}^\top)\boldsymbol{s} - \boldsymbol{u}^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{r} = 0, \tag{10}$$

and

$$(\boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top)(\boldsymbol{s} - \boldsymbol{r}) = \boldsymbol{0}. \tag{11}$$

Determine the maximum number of positive entries in $\boldsymbol{r}$ and $\boldsymbol{s}$ satisfying (10) and (11). Positive entries in $\boldsymbol{r}$ and $\boldsymbol{s}$ indicate fitted all probabilities fixed by the conditioning event

$$\{\boldsymbol{Y} | \boldsymbol{Z}^\top \boldsymbol{Y} = \boldsymbol{u}\}, \tag{12}$$

at 0 and 1 respectively. Then $\boldsymbol{Y} | \boldsymbol{Z}^\top \boldsymbol{Y} = \boldsymbol{u}$ remains the same if observations with positive value for either $\boldsymbol{r}$ or $\boldsymbol{s}$ are removed prior to calculation.

Here $\boldsymbol{s}$ and $\boldsymbol{r}$ represent potential deviations of the positive and negative parts of logit of observation probabilities moving to infinity, respectively. Equation (10) forces that directions towards positive infinity are blocked for observations with any negative results, and that directions toward negative infinity are blocked for observations with any positive results. Equation (11) limits directions of infinite estimates to the range of the design matrix.

Additional steps are necessary for inference in logistic regression when parameters not of interest are estimated at infinity (Kolassa, 1997). Partition $\boldsymbol{Z}$ as $(\boldsymbol{Z}_i, \boldsymbol{Z}_n)$. Here $\boldsymbol{Z}_i$ is the matrix of covariates associated with interest parameters, and $\boldsymbol{Z}_n$ is the matrix of covariates associated with other ("nuisance") parameters. Obtain $\boldsymbol{r}$ and $\boldsymbol{s}$ from Theorem 1 using the reduced design matrix $\boldsymbol{Z}_n$, let $\boldsymbol{Z}_{n1}$ and $\boldsymbol{Z}_{i1}$ be the matrices $\boldsymbol{Z}_n$ and $\boldsymbol{Z}_i$, respectively, with rows corresponding to

nonzero entries in $\boldsymbol{r}$ or $\boldsymbol{s}$ omitted, let $\boldsymbol{Z}_{n2}$ be the matrix $\boldsymbol{Z}_n$ with rows corresponding to nonzero entries in $\boldsymbol{r}$ or $\boldsymbol{s}$ retained, and let $\boldsymbol{Y}_1$ be the vector $\boldsymbol{Y}$ with entries corresponding to nonzero entries in $\boldsymbol{r}$ or $\boldsymbol{s}$ omitted. Let $\hat{\boldsymbol{Y}}_2$ be a vector with length equal to the number of nonzero entries in $\boldsymbol{r}$ or $\boldsymbol{s}$, with entries 1 corresponding to positive entries of $\boldsymbol{s}$, and 0 corresponding to positive entries for $\boldsymbol{r}$. Then the conditional distributions $\boldsymbol{Z}_i^\top \boldsymbol{Y} | \boldsymbol{Z}_n^\top \boldsymbol{Y}$ and $\boldsymbol{Z}_{n1}^\top \boldsymbol{Y}_1 + \boldsymbol{Z}_{n2} \hat{\boldsymbol{Y}}_2 | \boldsymbol{Z}_{n1}^\top \boldsymbol{Y}_1$ are the same.

The same phenomenon of likelihood monotonicity occurs in multinomial regression; this problem may be addressed by constructing design matrix (6), with as many rows as there are alternatives in the original multinomial regression model, and with $\mathcal{K} = \mathcal{D} + \mathcal{P}$ columns, formed by combining the original covariate vectors with new indicator variables (Kolassa, 2016). Theorem 1 is then used to indicate which subjects in the logistic regression should be omitted to leave conditional inference for the variables of interest unchanged, and to yield finite conditional maximum likelihood estimators. Since the partial likelihood for proportional hazards regression is the same as the likelihood for the multinomial regression, this same technique is applied to the multinomial regression configuration as described in Section 3.3.

5 | THE PROPOSED METHOD

We propose constructing hypothesis tests and confidence intervals for interest parameters in proportional hazards regressions models in which parameters not of direct interest are estimated at infinity, by converting the dataset to one yielding approximately equivalent inference and having a likelihood that is nonmonotone in all contrasts of parameters.

First, construct a multinomial regression data set from the proportional hazards dataset. The new multinomial regression dataset has as many trials $\mathcal{P}$ as there are event times. As in Section 3.3, for trial m, the risk set $\mathcal{A}_m$ is the set of observations whose event and censoring times are greater than or equal to that under consideration, W_m, and h_i is the number of subjects in the risk set $\mathcal{A}_m$. The associated alternative covariate vectors are the covariate vectors for these individuals. In the multinomial regression model of Section 3.1, no relationship need exist among alternatives in the various trials; our proposal is to reuse covariate vectors until the individuals are removed from the risk set $\mathcal{A}_m$. This identification of the multinomial and proportional hazards regression model is motivated by noting that likelihood contributions multiply for the multinomial model, because of independence, and for the proportional hazards model, because they are defined conditionally. Ideally, frequentist inference for these two models will be different, because they arise from different sampling distributions; we substitute multinomial regression for proportional hazards regression for the purpose of eliminating contrasts estimated at infinity (Kolassa, 2016).

We express multinomial regression as conditional logistic regression, and analyze the conditional logistic regression to identify subjects whose fitted probabilities are 0 or 1. This determines which multinomial subjects have selection probability either 0 or 1, under the conditioning event, and, in turn, implies which survival subjects are guaranteed either to have or to fail to have the next event. We remove these from the proportional hazards regression.

After the proportional hazards likelihood is converted to a multinomial regression likelihood, and multinomial observations indicated by nonzero components of $\boldsymbol{r}$ or $\boldsymbol{s}$ from the algorithm of Theorem 1 are removed, the multinomial regression likelihood containing the remaining observations is maximized.

It may happen that all contributions from a subject in the original proportional hazards regression are removed at every time in this manner, but generally the resulting multinomial regression data set will have subjects present at some times but not at others.

Maximizers of the likelihood of the reduced multinomial regression model are finite, and standard normal theory, or higher order, inference can be applied.

Without loss of generality, assume that inference on the first covariate is desired. If ℓ is as in (5), with observations determined as above removed, and with covariates making the covariate matrix no longer of full rank removed, let $\hat{\boldsymbol{\beta}}$ be the maximizer of $\ell(\boldsymbol{\beta})$, and let $\tilde{\boldsymbol{\beta}}$ represent the maximizer of $\ell(\boldsymbol{\beta})$ subject to $\beta_1 = 0$. Let $\hat{\sigma}_1^2$ represent the inverse of the first diagonal entry of $\ell''(\hat{\boldsymbol{\beta}})^{-1}$. Then the Wald statistic for testing the null hypothesis $\beta_1 = 0$ is $\hat{z} = \hat{\beta}_1/\hat{\sigma}_1$, and the signed root of the likelihood ratio statistic is $\hat{w} = \sqrt{2[\ell(\hat{\boldsymbol{\beta}}) - \ell(\tilde{\boldsymbol{\beta}})]}$; both of these statistics have a null hypothesis distribution that is approximately standard normal.

The proposed approach might be extended to hypotheses involving multiple parameters simultaneously. As before, covariates of interest are removed from the design matrix, the proportional hazards regression is transformed in to the associated multinomial regression, and then into conditional logistic regression. The appropriate subjects are removed, and then standard profile likelihood methods, including Wald and likelihood ratio statistics. are employed.

It is useful to develop some intuition behind this procedure. Suppose that the likelihood for the model with the interest parameter removed is monotone. Likelihood monotonicity implies that there is a $\boldsymbol{\beta}^*$ such that $L(\lambda\boldsymbol{\beta}^*)$ is increasing in λ, and this monotonicity occurring with the interest parameter removed implies that such a $\boldsymbol{\beta}^*$ can be chosen with the component associated with the interest parameter being zero. In the multinomial formulation of the proportional hazard partial likelihood, the observations are indexed by an event time m and a subject $j \in \mathcal{A}_m$ at risk at that time. Consider the limiting value for the probability associated with such a probability, $\lim_{\lambda\to\infty} \pi_{mj}(\lambda\boldsymbol{\beta}^*)$. When $\boldsymbol{x}_{mj}\boldsymbol{\beta}^* > 0$, and $\boldsymbol{x}_{mk}\boldsymbol{\beta}^* \leq 0$ for all $k \in \mathcal{A}_m, k \neq j$, then $\lim_{\lambda\to\infty} \pi_{mj}(\lambda\boldsymbol{\beta}^*) = 1$. When $\boldsymbol{x}_{mj}\boldsymbol{\beta}^* < 0$, or when $\boldsymbol{x}_{mj}\boldsymbol{\beta}^* = 0$ and $\boldsymbol{x}_{mk}\boldsymbol{\beta}^* > 0$ for some $k \in \mathcal{A}_m$, then $\lim_{\lambda\to\infty} \pi_{mj}(\lambda\boldsymbol{\beta}^*) = 0$. These cases are indicated by checking for positive values of $\boldsymbol{r}$ or $\boldsymbol{s}$ of Theorem 1. The set of observations having a positive component of $\boldsymbol{r}$ or $\boldsymbol{s}$ are the same as the components of $\boldsymbol{Y}$ whose value is fixed by the conditioning event (12) (Kolassa, 1997).

A related technique allows for detecting the presence of parameters estimated at infinity without providing for inference after detection (Clarkson & Jennrich, 2000).

5.1 | An illustration of the method

Consider a simple example with seven observations and two covariates, X and W, where covariate X will be of main interest. The dataset is given in Table 4. Expanding this table into the associated multinomial regression model gives two multinomial observations, generated by selecting subject 3 having the event at time 1 from among those at risk (all seven observations) at time 1, and by selecting subject 2 to have the event at time 3 from among those (1, 2, and 7) at risk at time 3. Even though subject 1 has an event at time 4, this does not add an additional multinomial trial, since subject 1 is the only subject at risk at time 4. One might express this multinomial dataset as in Table 5. The final two columns of Table 5 complete the corresponding logistic regression data set; the multinomial regression is equivalent to the logistic regression on the variables X and W, and the two indicator functions, without an intercept, and conditioned on the sufficient statistics for the last two columns. The values of these statistic in this example are the sums of products of the column Event and each of the last two columns; these two values are both 1, representing one

TABLE 4 Small example

Observation	Time	Event indicator	*X*	*W*
1	4	1	0	0
2	3	1	2	0
3	1	1	1	0
4	1	0	1	0
5	2	0	1	1
6	2	0	0	1
7	3	0	0	1

TABLE 5 Derived small multinomial dataset

Time	Subject	Event	Time	*X*	*W*	Time 1 indicator	Time 3 indicator
1	1	0	4	0	0	1	0
1	2	0	3	2	0	1	0
1	3	1	1	1	0	1	0
1	4	0	1	1	0	1	0
1	5	0	2	1	1	1	0
1	6	0	2	0	1	1	0
1	7	0	3	0	1	1	0
3	1	0	4	0	0	0	1
3	2	1	3	2	0	0	1
3	7	0	3	0	1	0	1

event at each event time. Inference on the parameter associated with X without consideration of the parameter associated with W requires conditioning on the sum of the product of the Event column and the W column. This sum is zero, and so by inspection, conditional on this event, subject 7 cannot have the event at time 3, and none of subjects 5, 6, or 7 can have the event at time 1. Hence subjects 5 and 6 are removed at time 1, and subject 7 is removed at times 1 and 3. The resulting multinomial regression fails to be of full rank, and so W is removed from the model as having an unidentifiable coefficient. The resulting reduced multinomial likelihood is maximized to generate the parameter estimate for X. The resulting parameter estimate for X is 0.9666, with standard error 0.9555 and p-value.312.

6 | EXAMPLES AND SIMULATIONS

We present the results of the proposed method applied to the breast cancer data, and then present the results of a simulation made under conditions like those of this data set, treating the variable T (representing stage) as the covariate of interest.

After converting the proportional hazards regression task into a conditional logistic regression task whose likelihood is exactly the same as the original partial likelihood, identifying

TABLE 6 Results of proportional hazards regression for breast cancer survival, after removing individuals with deterministic selection probabilities

Variable	Estimate	SE	*p*
T	1.4129	0.5034	.0050
N	1.0329	0.4284	.0159
G	NA	NA	NA
CD	0.4436	0.4468	.3208

observations whose selection probabilities are zero or one based on covariates not of interest, removing the contributions of indicated subjects at the indicated times, and performing multinomial regression on the reduced dataset provides Table 6. Results were calculated using an R package (Kolassa & Zhang, 2021). The effect of *T* differs in Tables 1,2, and 6. The result of the proposed method differs substantially from the failed-to-converge result (1.413 vs. 1.650), and from the result dropping offending covariate (1.413 vs. 1.279). One covariate, *G*, was removed because of non-identifiability.

6.1 | Simulations from models similar to that of the motivating example

One should check that the proposed procedure does not adversely effect the distribution of *p*-values under the null hypothesis.

In order to assess the true level of tests constructed from nominal test constructions, datasets from a model consistent with the motivating example. First, a Weibull regression model was fit to the dataset. The parameter associated with *T* was set to zero. Then, 1,000 new survival times were simulated using these fitted parameters. and the observations were censored at 72 months. Since some of these regression parameters are quite large, many of the simulated datasets have infinite estimates. Tables summarizing simulation results also report proportions of data sets with infinite estimates. Wald and likelihood ratio *p*-values for the test of the null hypothesis that the coefficient associated with *T* is zero, for both the proposed method and the regularization method (Heinze & Schemper, 2001), were calculated, and their empirical distributions are plotted in Figure 1. In this case, the two penalized methods and the proposed method with the Wald statistic fail to control Type I error; the proposed method using the likelihood ratio statistic controls Type 1 error, and in fact has a true level close to the nominal level throughout. This pattern is not universal; Figure 2 shows results for the variable CD. In this figure, none of the four procedures properly control Type I error, and none is particularly worse than the others.

6.2 | Special cases

After reducing the terms in the partial likelihood, consequently removing redundant noninterest covariates, and adding the interest parameter back in, the estimate of the interest parameter may still be infinite. Again, placing the interest parameter as the first parameter, then $\hat{\beta}_1$ is one of $\pm\infty$. In this case, the Wald test statistic is not defined. In subsequent calculations, the Wald *p*-value is set to zero. A preferred approach involves a continuity correction, as was done for

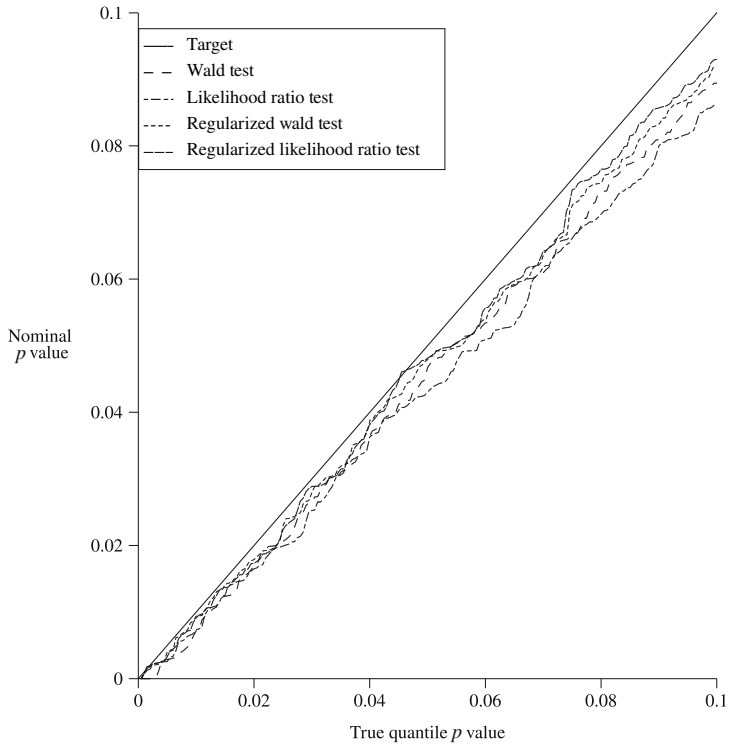

FIGURE 1 *p*-Value comparisons for simulated breast cancer data, Variable *T*

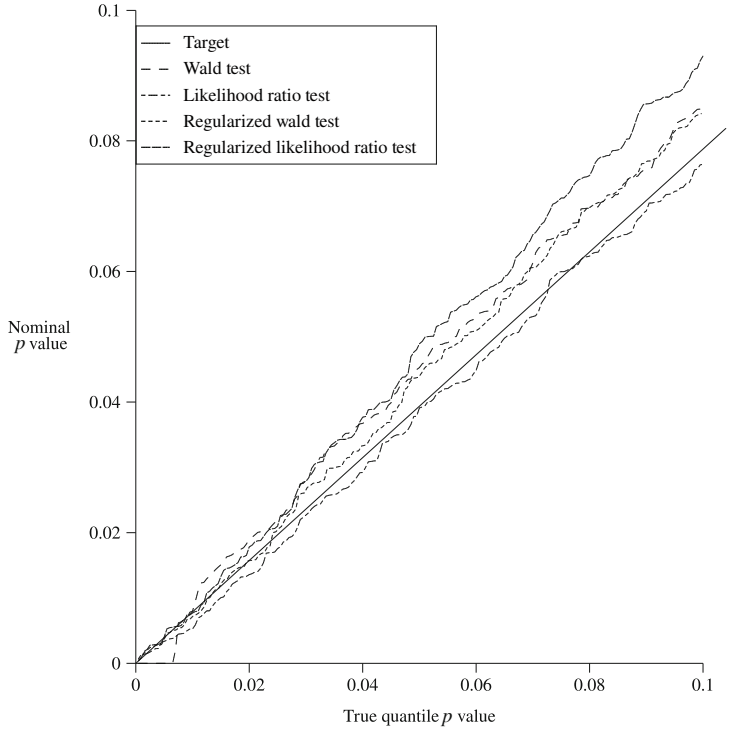

FIGURE 2 *p*-Value comparisons for simulated breast cancer data, Variable CD

logistic regression in §2, but in this case continuity correct involves not simply shifting U_1 from (8), since the second term in (7) is also data-dependent.

In some extreme cases, application of Theorem 1 removes all observations from the regression without the interest parameter; no observations are left to estimate or test the interest parameter. In the analogous cases involving conditioning in a canonical exponential family, no inferential statement can be made. By analogy, we do not recommend inference in such a case.

6.3 | Simulations to examine operating characteristics of the proposed procedure

The regularization approach to this issue was assessed using a simulation study to investigate the operating characteristics of their procedure, investigating simulated datasets involving various numbers of observations, various numbers of dichotomous covariates, various relative risks, various censoring proportions, and various degrees of imbalance in the covariates (Heinze & Schemper, 2001).

Tables 7 and 8 incorporate a subset of these simulation settings. Attention was restricted to the sample size 50, and the censoring proportion was restricted to less than or equal to 0.5. For each simulation setting, 1,000 datasets are simulated. Each simulated data set consists of 50 observations independently sampled from an exponential distribution. Random indicators of censoring were simulated, with proportion c censored. In each case, five dichotomous covariates were simulated, with proportion $B/(B+1)$ taking the value 0, and the remainder taking the value 1. Data were simulated with various relative risks R for each parameter. Relative risks are set to 1, 2, or 4. Larger censoring probabilities produced substantial numbers of simulated cases in which the step removing observations to account for the effect of parameters not of interest lead to empty models, as described in Section 6.2, and, furthermore, the standard implementation of the regularization

TABLE 7 Level and power for monotone likelihood methods

Covariate balance	$B=1$	$B=1$	$B=4$	$B=4$
Censoring proportion	$c=0$	$c=0.5$	$c=0$	$c=0.5$
Proposed Wald level $R=1$	0.0810	0.0660	0.0820	0.0800
HS Wald level $R=1$	0.0831	0.0691	0.0870	0.0840
Proposed LR level $R=1$	0.08200	0.0720	0.0870	0.0780
HS LR level $R=1$	0.08200	0.0650	0.0880	0.0810
Proposed Wald power $R=2$	0.6750	0.3840	0.5590	0.3540
HS Wald power $R=2$	0.6787	0.4010	0.5690	0.3660
Proposed LR power $R=2$	0.6820	0.4060	0.5440	0.3180
HS LR power $R=2$	0.6730	0.3800	0.5720	0.3550
Proposed Wald power $R=4$	0.98800	0.8530	0.9550	0.7010
HS Wald power $R=4$	0.98800	0.8580	0.9590	0.7080
Proposed LR power $R=4$	0.98800	0.8640	0.9510	0.6620
HS LR power $R=4$	0.98800	0.8430	0.9600	0.6820

Note: HS denotes the regularization approach (Heinze & Schemper, 2001), and LR represents likelihood ratio.

TABLE 8 Median bias for monotone likelihood methods

Covariate balance	$B=1$	$B=1$	$B=4$	$B=4$
Censoring proportion	**c = 0**	**c = 0.5**	**c = 0**	**c = 0.5**
Proposed $R=1$	−0.0057	0.0106	0.0394	0.0062
HS $R=1$	−0.0043	0.0096	0.0710	0.0771
Proposed $R=2$	0.0584	0.0655	0.0839	0.0823
HS $R=2$	0.0489	0.0449	0.1053	0.1159
Proposed $R=4$	0.0858	0.0911	0.1446	0.1288
HS $R=4$	0.0548	0.0321	0.1468	0.1439

Note: HS denotes the regularization approach (Heinze & Schemper, 2001), and LR represents likelihood ratio.

approach (Heinze & Ploner, 2018) failed to converge in many of these cases, even after changing convergence parameters as directed by package suggestions.

The first four lines of Table 7 show that both the proposed method and the regularization method produce asymptotic tests of nominal level 0.05 with true level concerningly far from 0.05, although in all cases the proposed method is closer to the target than is the regularization method. The lines with relative risk R not equal to 1, representing datasets generated with true parameters all either log(2) or log(4), represent power. The critical values for these tests are taken from the 0.05 quantile of the 1,000 test statistic values simulated under the null hypothesis, with Wald and likelihood ratio statistics treated separately. Powers for the proposed and regularization tests are comparable.

Table 8 shows median bias for the first parameter. While estimation has not been the key aim of this paper, the median bias was generally better for the regularized method for balanced covariates, and the proposed method performed better for imbalanced covariates.

Table 9 shows test operating characteristics in cases with more censoring, and larger effect sizes. These situations more often generate monotone partial likelihood (Pagui & Colosimo, 2020). Simulations are as above, except that the interest variable is taken as standard normal rather than dichotomous.

Again, the proposed approach generally performs better than regularization using Jefferies prior; in some cases this improvement is substantial.

One might apply the method of Section 2.1 to provide a confidence interval for the coefficient of G in the breast cancer dataset, by applying (2) and (3) to the statistic $\boldsymbol{U}$ of (8). As noted in Section 2.1, one of these continuity-corrected statistics is associated with an infinite proportional hazards estimate, and so an infinite confidence interval end point; the other is associated with a convergent estimator, and gives the finite confidence interval endpoint. In this case, the resulting 95% confidence interval is $(-0.514, \infty)$. In contrast, the penalized method (Heinze & Schemper, 2001) gives the interval (1.466, 1451.945); the finite upper bound is entirely an artifact of the penalty.

As is typical for p-value or confidence interval calculations involving discrete data, the sufficient statistic associated with the parameter of interest must be corrected for continuity before applying a Gaussian approximation. Standard implementations of Cox regression rely on a sample size large enough to make a continuity correction irrelevant; infinite estimates only occur when sample sizes are small enough to make continuity correction essential.

TABLE 9 Level and power for monotone likelihood methods, higher censoring case

Covariate balance	$B=1$	$B=1$	$B=4$	$B=4$
Censoring proportion	$c=0.75$	$c=0.90$	$c=0.75$	$c=0.90$
Proposed Wald level $R=1$	0.066000	0.05600	0.06300	0.05200
HS Wald level $R=1$	0.066000	0.05906	0.06400	0.06006
Proposed LR level $R=1$	0.07100	0.06700	0.06800	0.05900
HS LR level $R=1$	0.06700	0.06000	0.06400	0.05800
Proposed Wald power $R=4$	0.998000	0.82100	0.99600	0.83400
HS Wald power $R=4$	0.998000	0.83868	0.99700	0.83267
Proposed LR power $R=4$	0.99800	0.86700	0.99800	0.86100
HS LR Power R=4	0.99700	0.82900	0.99600	0.82300
Proposed Wald level $R=16$	1.000000	0.95300	1.00000	0.96000
HS Wald level R=16	1.000000	0.96389	0.99900	0.95377
Proposed LR level $R=16$	1.000000	0.97200	1.00000	0.98300
HS LR level $R=16$	1.000000	0.95800	0.99900	0.94900

While the computations performed by this package are complex and time-consuming, the user interface is only minimally more complicated than that of the standard or penalized Cox regression. For example, in order to apply the proposed method for the breast cancer data included in the `coxphf` library, and obtain Table 6, one need only do

```
library(PHInfiniteEstimates)
data(breast) # From library coxphf
bcfit<-coxph(Surv(TIME,CENS)~ T+ N+ G+ CD,data=breast,x=TRUE)
fixcoxph(bcfit,bcfit$x,"T",Surv(TIME,CENS)~ T+ N+ G+ CD)
```

All calculations in this manuscript were performed using R and the package `PHInfiniteEstimates`. No original data was used in this manuscript.

7 | CONCLUSION

This paper presents a method for inference in proportional hazards in the presence of likelihood monotonicity which applies a known solution to this problem for the analogous multinomial regression likelihood. The chief existing method addressing this problem is regularization of the partial likelihood using the Jeffreys prior. Both procedures have similar operating characteristics. The proposed method has advantages over the existing method, in that it reduces to the standard inferential approach in the absence of likelihood monotonicity, and, after removing monotonicity, represents an easier numerical optimization problem. Furthermore, the proposed method avoids the strong influence on parameter estimation exhibited by the particular regularization prior applied (Almeida, Colosimo, & Mayrink, 2018). Its disadvantages are the computational complexity of the steps necessary to remove monotonicity.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in CRAN at https://cran.r-project.org/. These data were derived from the following resources available in the public domain: - Package, https://cran.r-project.org/.

ORCID

John E. Kolassa https://orcid.org/0000-0002-8246-4276

REFERENCES

Almeida, F. M., Colosimo, E. A., & Mayrink, V. D. (2018). Prior specifications to handle the monotone likelihood problem in the cox regression model. *Statistics and Its Interface*, *11*, 687–698.

Clarkson, D. B., & Jennrich, R. I. (2000). *Computing extended maximum likelihood estimates for cox proportional-hazards models*. In F. T. Bruss & L. Le Cam (Eds.), *Game theory, optimal stopping, probability and statistics Lecture notes–monograph series* (Vol. *35*, pp. 205–217). Beachwood, OH: Institute of Mathematical Statistics Institute of Mathematical Statistics.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B: Methodological*, *34*(2), 187–220.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*(1), 27–38.

Heinze, G., & Ploner, M. (2003). Fixing the nonconvergence bug in logistic regression with SPLUS and SAS. *Computer Methods and Programs in Biomedicine*, *71*(2), 181–187.

Heinze, G., & Ploner, M. (2018). *coxphf: Cox regression with Firth's penalized likelihood. Comprehensive R archive network. R package version 1.13*. Medizinische Universität Wein.

Heinze, G., & Puhr, R. (2010). Bias-reduced and separation-proof conditional logistic regression with small or sparse data sets. *Statistics in Medicine*, *29*(7-8), 770–777.

Heinze, G., & Schemper, M. (2001). A solution to the problem of monotone likelihood in cox regression. *Biometrics*, *57*(1), 114–119.

Hjort, N. L. (1986). Bayes estimators and asymptotic efficiency in parametric counting process models. *Scandinavian Journal of Statistics*, *13*(2), 63–85.

Kolassa, J. E. (1997). Infinite parameter estimates in logistic regression, with application to approximate conditional inference. *Scandinavian Journal of Statistics*, *24*(4), 523–530.

Kolassa, J. E. (2016). Inference in the presence of likelihood monotonicity for polytomous and logistic regression. *Advances in Pure Mathematics*, *6*(5), 331–341.

Kolassa, J. E., & Zhang, J. (2021). *PH infinite estimates: Tools for inference in the presence of a monotone likelihood*. Comprehensive R Archive Network. R package version 1.8, Rutgers, the State University of New Jersey.

Lösch, A., Tempfer, C., Kohlberger, P., Joura, E. A., Denk, M., Zajic, B., … Kainz, C. (1998). Prognostic value of cathepsin d expression and association with histomorphological subtypes in breast cancer. *British Journal of Cancer*, *78*(2), 205–209 9683294[pmid].

Mansournia, M. A., Geroldinger, A., Greenland, S., & Heinze, G. (2017). Separation in logistic regression: Causes, consequences, and control. *American Journal of Epidemiology*, *187*(4), 864–869.

Pagui, E. C. K., & Colosimo, E. A. (2020). Adjusted score functions for monotone likelihood in the cox regression model. *Statistics in Medicine*, *39*, 1558–1572.

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

SAS Institute, Inc. (2019). *SAS/STAT(R) 15.1 user's guide, Proc PHREG model statement*. Cary, NC: SAS Institute, Inc.

Therneau, T. M. (2015). *A package for survival analysis in S. Comprehensive R archive network. Version 2.38*, Vienna, Austria.

Wu, J., de Castro, M., Schifano, E. D., & Chen, M.-H. (2018). Assessing covariate effects using jeffreys-type prior in the cox model in the presence of a monotone partial likelihood. *Journal of Statistical Theory and Practice*, *12*(1), 23–41.

Zhang, J., & Kolassa, J. (2013). A practical procedure to find matching priors for frequentist inference. *Communication in Statistics - Theory and Methods*, *42*, 2758–2767.

How to cite this article: Kolassa, J. E., & Zhang, J. (2023). Inference in the presence of likelihood monotonicity for proportional hazards regression. *Statistica Neerlandica*, 1–18. https://doi.org/10.1111/stan.12287