

# surrosurv: an R Package for the Evaluation of Failure Time Surrogate Endpoints in Individual Patient Data Meta-Analyses of Randomized Clinical Trials

Federico Rotolo<sup>a,b,\*</sup>, Xavier Paoletti<sup>a,b</sup>, Stefan Michiels<sup>a,b</sup>

<sup>a</sup>Service de Biostatistique et d'Épidémiologie, Gustave Roussy, Université Paris-Saclay. 114, Rue Edouard Vaillant, Villejuif 94805 France  
<sup>b</sup>INSERM U1018 OncoStat, CESP, Université Paris-Sud, Université Paris-Saclay

---

## Abstract

*Background and Objective.* Surrogate endpoints are attractive for use in clinical trials instead of well-established endpoints because of practical convenience. To validate a surrogate endpoint, two important measures can be estimated in a meta-analytic context when individual patient data are available: the  $R_{\text{indiv}}^2$  or the Kendall's  $\tau$  at the individual level, and the  $R_{\text{trial}}^2$  at the trial level. We aimed at providing an R implementation of classical and well-established as well as more recent statistical methods for surrogacy assessment with failure time endpoints. We also intended incorporating utilities for model checking and visualization and data generating methods described in the literature to date.

*Methods.* In the case of failure time endpoints, the classical approach is based on two steps. First, a Kendall's  $\tau$  is estimated as measure of individual level surrogacy using a copula model. Then, the  $R_{\text{trial}}^2$  is computed via a linear regression of the estimated treatment effects; at this second step, the estimation uncertainty can be accounted for via measurement-error model or via weights. In addition to the classical approach, we recently developed an approach based on bivariate auxiliary Poisson models with individual random effects to measure the Kendall's  $\tau$  and treatment-by-trial interactions to measure the  $R_{\text{trial}}^2$ . The most common data simulation models described in the literature are based on: copula models, mixed proportional hazard models, and mixture of half-normal and exponential random variables.

*Results.* The R package `surrosurv` implements the classical two-step method with Clayton, Plackett, and Hougaard copulas. It also allows to optionally adjusting the second-step linear regression for measurement-error. The mixed Poisson approach is implemented with different reduced models in addition to the full model. We present the package functions for estimating the surrogacy models, for checking their convergence, for performing leave-one-trial-out cross-validation, and for plotting the results. We illustrate their use in practice on individual patient data from a meta-analysis of 4069 patients with advanced gastric cancer from 20 trials of chemotherapy.

*Conclusions.* The `surrosurv` package provides an R implementation of classical and recent statistical methods for surrogacy assessment of failure time endpoints. Flexible simulation functions are available to generate data according to the methods described in the literature.

*Keywords:* surrogate endpoints, survival, individual patient data meta-analyses, R, copula, mixed models

---

## 1. Introduction

Surrogate endpoints are endpoints which can reliably be used instead of well-established (true) endpoints and which yield improved practical convenience in terms of lower cost, more rapid occurrence, increased ease of assessment, or reduced invasiveness [1]. Two conditions must be fulfilled for surrogate endpoint to be reliable: it must be strongly associated with the true endpoint at the individual level and the effect of the treatment on it must be strongly associated with the effect on the true endpoint. In

a meta-analytic context and when the endpoints are gaussian [2], the usual measure of individual level surrogacy is the  $R_{\text{indiv}}^2$  between the *endpoints*, which measures the part of variability of the true endpoint  $T$  explained by the surrogate endpoint  $S$ . At the trial level, the usual measure of surrogacy is given by the  $R_{\text{trial}}^2$  between the *treatment effects* on the two endpoints, that measures the part of variability of the *treatment effect* on  $T$  explained by the *treatment effect* on  $S$ .

In the case of failure time (survival) endpoints, the classical methods developed for normally-distributed endpoints cannot be used because of right censoring. Burzykowski and colleagues [3] developed a meta-analytic model for failure time endpoints that measures individual level surro-

---

\*Corresponding author

Email address: federico.rotolo@gustaveroussy.fr (Federico Rotolo)

gacy in terms of Kendall's  $\tau$  [4] and trial level surrogacy in terms of  $R_{\text{trial}}^2$ . This method is largely employed in numerous applications in the medical literature. Because of some limitations including convergence issues, the interpretation of the results is difficult in some cases [5, 6]. There exists a well-known connection between the proportional hazard model and the Poisson log-linear model, which ensures that estimates of the regression parameters are equivalent in the two models, upon appropriate parameterization [7, 8]. Recently, we used this connection in the bivariate context of surrogate endpoint validation, by proposing a joint model for the two treatment effects adjusted for individual dependence and baseline heterogeneity across trials [9]. We used bivariate mixed proportional hazard models [10], which are the most natural adaptation of the above-mentioned meta-analytic approach by Buyse et al. [2] to the survival case. Indeed, the original approach for gaussian data consisted of a single model, incorporating random effects for both the individual and the trial level.

In the present paper, we show how the classical and more recent models can be fitted by use of the R [11] package `surrosurv` [12]. The package provides utilities for convergence assessment and tools for leave-one-trial-out cross-validation, which allow contrasting the observed treatment effect in each trial to those predicted from the others. User-friendly functions allow the user to clearly show the results of the estimated models. We illustrate the available functions using individual data of a meta-analysis of 20 randomized trials of chemotherapy, including 4069 patients with advanced/recurrent gastric cancer [13, 14].

## 2. Computational methods and theory

Let  $T_{ij}$  and  $S_{ij}$  be the times to the true and the surrogate endpoints, respectively, for patient  $j \in \{1, \dots, n_i\}$  in trial  $i \in \{1, \dots, N\}$ . Let  $Z_{ij}$  be the indicator of the treatment arm to which the  $j$ -th patient in the  $i$ -th trial has been randomized.

### 2.1. Two-step copula approach

The model proposed by Burzykowski et al. [3] for failure time endpoints consists in two steps, one for the individual and one for the trial level.

*Individual-level.* At the first step, the authors defined the bivariate proportional hazard model by means of the marginal hazard functions and of the copula function to account for their dependence:

$$\begin{cases} h_{S_{ij}}(s; Z_{ij}) = h_{S_i}(s) \exp\{\alpha_i Z_{ij}\} \\ h_{T_{ij}}(t; Z_{ij}) = h_{T_i}(t) \exp\{\beta_i Z_{ij}\} \\ C_\theta(S_{S_{ij}}(s), S_{T_{ij}}(t)) \end{cases} \quad (1)$$

where  $h_{S_i}(s)$  and  $h_{T_i}(s)$  are the trial-specific baseline hazards,  $\alpha_i$  and  $\beta_i$  the treatment effects, and  $S_{S_{ij}}(s)$  and

$S_{T_{ij}}(t)$  the survival functions associated to  $h_{S_{ij}}$  and  $h_{T_{ij}}$ . The dependence parameter  $\theta$  is reparametrized into the individual-level Kendall's  $\tau$ , according to the copula function thanks to the `tau()` function in the R package `copula` [15, 16].

In the `surrosurv` package, Weibull marginal hazards are implemented, together with three copula functions:

- the Clayton copula [17]

$$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \quad (2)$$

with  $\theta > 0$  and Kendall's  $\tau = \theta/(\theta + 2)$ ;

- the Plackett copula [18]

$$\begin{aligned} C_\theta(u, v) &= [Q - R^{1/2}]/[2(\theta - 1)], \\ Q &= 1 + (\theta - 1)(u + v), \\ R &= Q^2 - 4\theta(\theta - 1)uv, \end{aligned} \quad (3)$$

with  $\theta > 0$  and Kendall's  $\tau$  computed using numerical integration as no analytical expression is available;

- the Hougaard copula [19]

$$C_\theta(u, v) = \exp\left(-\left[(-\ln u)^{1/\theta} + (-\ln v)^{1/\theta}\right]^\theta\right), \quad (4)$$

with  $\theta \in (0, 1)$  and Kendall's  $\tau = 1 - \theta$ .

Further details on these three copula models can be found in the vignette(`'copula'`, `package = 'surrosurv'`).

*Trial level.* At the second step, the estimates of the treatment effects obtained at the first step are assumed to follow the mixed model

$$\begin{pmatrix} \hat{\alpha}_i \\ \hat{\beta}_i \end{pmatrix} = \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} + \begin{pmatrix} \epsilon_{ai} \\ \epsilon_{bi} \end{pmatrix}, \quad (5)$$

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \mathbf{D}\right) \quad (6)$$

$$\mathbf{D} = \begin{pmatrix} d_a^2 & d_a d_b \rho_{\text{trial}} \\ d_a d_b \rho_{\text{trial}} & d_b^2 \end{pmatrix}, \quad (7)$$

$$\begin{pmatrix} \epsilon_{ai} \\ \epsilon_{bi} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{\Omega}_i\right) \quad (8)$$

$$\mathbf{\Omega}_i = \begin{pmatrix} \omega_{ai}^2 & \omega_{ai}\omega_{bi}\rho_{\epsilon i} \\ \omega_{ai}\omega_{bi}\rho_{\epsilon i} & \omega_{bi}^2 \end{pmatrix}. \quad (9)$$

where  $(\alpha_i, \beta_i)'$  are the true treatment effects and  $(\epsilon_{ai}, \epsilon_{bi})'$  the estimation errors.

The trial-level surrogacy measure is  $R_{\text{trial}}^2 = \rho_{\text{trial}}^2$ . In practice, we compute the  $\rho_{\text{trial}}$  via a linear regression of the  $\beta_i$ 's over the  $\alpha_i$ 's adjusted by measurement error by fixing the  $\mathbf{\Omega}_i$ 's at their estimates from the first step [20] by using the `mvmeta` package [21, 22]. This adjusted (for measurement error) model is sometimes computationally challenging and does not always converge. The `surrosurv` package

returns also the so-called unadjusted  $R_{\text{trial}}^2$ , obtained using a linear regression — equivalent to fixing all the elements of  $\mathbf{\Omega}_i$  equal to 0 — by weighting the observations  $(\alpha_i, \beta_i)'$  by the trial size, in order to account somehow indirectly and approximately for the standard error of the estimates.

## 2.2. One-step mixed Poisson approach

In a recent paper [9], we assumed that the bivariate proportional hazard model given by the first two lines of equation (1) holds conditionally on an individual random effect  $u_{ij} \sim \mathcal{N}(0, \sigma_{\text{indiv}}^2)$ :

$$\begin{cases} h_{Sij}(s | u_{ij}) = h_{Si}(s) \exp\{u_{ij} + \alpha_i Z_{ij}\} \\ h_{Tij}(t | u_{ij}) = h_{Ti}(t) \exp\{u_{ij} + \beta_i Z_{ij}\}. \end{cases} \quad (10)$$

Note that this corresponds to a shared frailty model with bivariate clusters [10]. The shared frailty term  $u_{ij}$  accounts for individual level dependence.

It is well-known (see for instance [7, 23]) that the parameters of Cox models can be estimated by fitting a so-called ‘auxiliary’ Poisson log-linear regression model, by dividing the time scale into intervals  $k = 1, \dots, K$ . The auxiliary Poisson model provides the same estimator as the Cox model if the bounds of the intervals are all the observed event times, and an approximation of the Cox estimators otherwise. In the surrogacy assessment context, the parameters of the bivariate frailty model (10) can be estimated via a bivariate mixed Poisson model

$$\begin{cases} \log\left(\mu_{Sij}^{(k)}\right) = \mu_{Si}^{(k)} + u_{ij} + \alpha_i Z_{ij} + \log\left(y_{Sij}^{(k)}\right) \\ \log\left(\mu_{Tij}^{(k)}\right) = \mu_{Ti}^{(k)} + u_{ij} + \beta_i Z_{ij} + \log\left(y_{Tij}^{(k)}\right) \end{cases} \quad (11)$$

with  $y_{Sj}^{(k)}$  and  $y_{Tj}^{(k)}$  the time spent at risk by subject  $i$  in trial  $j$  for each endpoint during the period  $k$ .

*Individual-level surrogacy.* The estimated variance of the shared frailties  $u_{ij}$  is  $\hat{\sigma}_{\text{indiv}}^2$  and can be used to estimate the Kendall’s  $\hat{\tau} = 4 \int_0^{\infty} s \mathcal{L}(s) \mathcal{L}^{(2)}(s) ds - 1$ , where  $\mathcal{L}(s)$  and  $\mathcal{L}^{(2)}(s)$  are the Laplace transform of the frailty distribution and its second derivative. As an analytic expression of  $\mathcal{L}(s)$  is not available for the log-normal frailty distribution, we approximated it using the Laplace method [24], implemented in the `fr.lognormal()` function in the `parfm` package [25, 26].

*Trial-level surrogacy.* In model (11), the trial-specific treatment effects are again assumed to follow the binormal distribution (7). Thus, the correlation  $\rho_{\text{trial}}$  between the two treatment effects provides us with the coefficient of determination  $R_{\text{trial}}^2 = \rho_{\text{trial}}^2$ , also referred to simply as  $R^2$ .

*Reduced Poisson models.* The `surrosurv` package can compute four reduced versions of the full model (11) that may turn out to be useful in case of convergence issues with the full model.

- Model **Poisson T** has random trial-treatment interactions  $\alpha_i$  and  $\beta_i$ , but does not incorporate individual effects ( $u_{ij} \equiv 0$ ). It assumes common baselines between trials ( $\mu_{Si}^{(k)} = \mu_S^{(k)}$ ,  $\mu_{Ti}^{(k)} = \mu_T^{(k)}$ ,  $\forall i$ ). This model provides only the trial-level measure of surrogacy  $R_{\text{trial}}^2$ .
- Model **Poisson I** contains individual random effects  $u_{ij}$ , but not the trial-specific treatment effects ( $\alpha_i = \alpha$ ,  $\beta_i = \beta$ ,  $\forall i$ ) and has common baselines between trials. This model provides only the individual-level measure of surrogacy  $\tau$ .
- Model **Poisson TI** incorporates both random trial-treatment interactions  $(\alpha_i, \beta_i)'$  and individual random effects  $u_{ij}$ , but still has common baselines between trials. It provides both individual-level and trial-level measures of surrogacy  $\tau$  and  $R_{\text{trial}}^2$ .
- Model **Poisson TIa** extends the model Poisson TI by accounting for trial-specific baseline risks, using shared random effects at the trial level:  $\mu_{Si} = \mu_S + m_i$ ,  $\mu_{Ti} = \mu_T + m_i$ , with  $m_i \sim \mathcal{N}(0, \sigma_m^2)$ .

## 3. Program description with a data example

We illustrate the use of the functions in the `surrosurv` package on the individual patient data of the advanced GASTRIC meta-analysis [13, 14].

```
> library(surrosurv)
Loading required package: optimx
> packageVersion('surrosurv')
[1] '1.1.24'
```

The individual data of the 4069 patients, already made public by [27], are also available directly in R in the `surrosurv` package:

```
> data('gastadv')
> nrow(gastadv)
[1] 4069
```

The data set contains the following variables:

```
> names(gastadv)
2 [1] "timeT" "statusT" "statusS"
4 [4] "timeS" "trialref" "trt" "id"
```

where `timeT` and `timeS` are the (possibly censored) times for overall survival (T) and for progression-free survival (S) expressed in days, `statusT` and `statusS` are the associated indicators of censoring (0) or event (1), `trialref` is the trial indicator ( $i$ ), `trt` is the treatment arm (−0.5 for control and 0.5 for chemotherapy), and `id` is the patient indicator ( $j$ ). Figure 1 shows the Kaplan–Meier curves for overall survival, the true endpoint  $T$ , and progression-free survival, the candidate surrogate  $S$ .

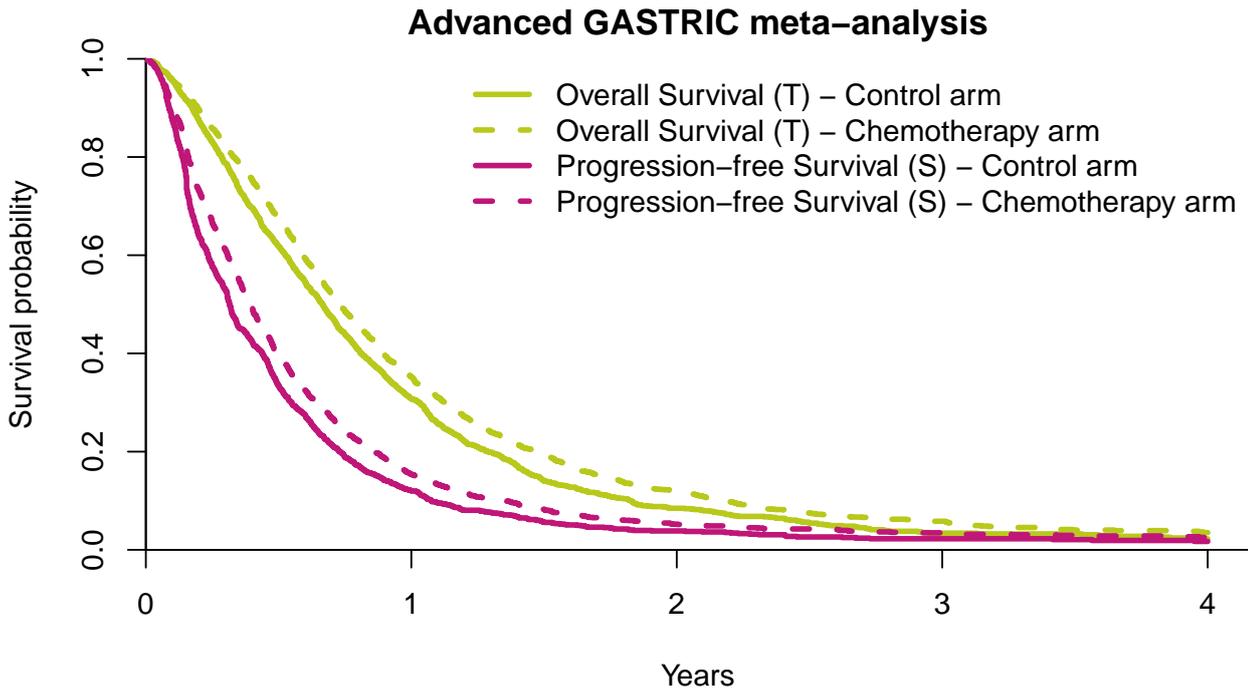


Figure 1: Kaplan–Meier curves for overall survival ( $T$ ) and progression-free survival ( $S$ ) in the advanced GASTRIC meta-analysis [13]

### 3.1. Fitting the surrogacy models

The surrogacy models presented in Section 2 can be fitted via the `surrosurv()` function.

The only mandatory argument for the `surrosurv()` function is `data`, which has to be a `data.frame` with columns

- `trialref`, a factor (with observations in at least two levels) containing the trial identifier;
- `trt`, the treatment arm, coded as  $-0.5$  vs.  $0.5$ ;
- `id`, a factor containing the patient id;
- `timeT` and `timeS`, two positive-valued numerical variables, containing the observed or censored times of the true endpoint  $T$  and of the candidate surrogate  $S$ , respectively;
- `statusT` and `statusS`, the censoring/event (0/1) indicators of  $T$  and  $S$ , respectively.

A second argument, `models`, can optionally contain the list of the models to fit (any of `clayton`, `plackett`, `hougaard`, `poisson`, `poissonT`, `poissonI`, `poissonTI`, or `poissonTIIa`). If not specified, all the models are fitted.

Two further parameters, `intWidth` and `nInts`, specify the width and the number of time intervals for data Poissonization. These parameters are passed to the function `poissonize()`, described in the Appendix (Sec. Appendix A). At most one of them can be specified. By default, `nInts = 8` which means that the study period is divided into eight intervals, the length of which is fixed so that  $1/8$ th of the observed events falls in each interval. This

choice is just based on our empirical experience, but we advise users to check results for different values.

The optimizer algorithm used for maximization of the likelihood can be chosen among those provided by the `optimx` package [28, 29]. To this end, the optional arguments `cop.OPTIMIZER` and `poi.OPTIMIZER` can be used for the copula models and the Poisson models, respectively.

The last parameter, `verbose`, is a logical value stating whether the function should print out the model being fitted (default: `FALSE`).

The surrogacy models for the advanced GASTRIC cancer meta-analysis are obtained as follows:

```

> allSurroRes <- surrosurv(gastadv, verbose
  = TRUE)
Computation may take very long. Please wait
...
- Estimating model: Clayton (5.3 mins)
- Estimating model: Plackett (4.8 mins)
5 - Estimating model: Hougaard (7.3 mins)
- Data poissonization (4.6 secs)
- Estimating model: Poisson T (1.2 mins)
- Estimating model: Poisson I (2.7 mins)
- Estimating model: Poisson TI (4 mins)
10 - Estimating model: Poisson TIIa (2 mins)
  
```

Note that the computation time of the surrogacy model estimation can be long. In this example, the computations required 27 mins on a PC with an Intel<sup>®</sup> quad-core CPU E3-1280 V2 with 3.60 GHz clock speed and 16GB of RAM. The results are an object of class `surrosurv` and the estimated Kendall's  $\tau$  and  $R^2$  can be easily displayed:

```

> allSurroRes
      kTau R2
Clayton unadj 0.61 0.45
Clayton adj   0.61 0.41
5 Plackett unadj 0.62 0.45
Plackett adj   0.62 0.4
Hougaard unadj 0.32 0.45
Hougaard adj   0.32 0.38
PoissonT       -.- 1
10 PoissonI     0.51 -.-
PoissonTI      0.51 0.63
PoissonTIIa    0.51 0.83

```

For each copula model, both the results with measurement error adjustment (adj) and without adjustment (unadj) are shown.

### 3.1.1. Assessing convergence

The function `convergence()` checks whether convergence criteria are met by each of the fitted models. Three convergence criteria are considered. The first criterion, `maxSgrad`, verifies whether the maximum gradient is small enough. The two other criteria, `minHev` and `minREev`, verify whether the minimum eigenvalue of the Hessian matrix of the fixed parameters (H) and of the covariance matrix of the random effects (RE) are big enough, in order to assure the positive definiteness of the two matrices. This check ensures that the estimated variances are positive. Two parameters can be used to tune the thresholds for ‘small enough’ maximum gradient and for ‘big enough’ minimum eigen value: `kkttol` ( $1e-2$  by default), and `kkt2tol` ( $1e-8$  by default).

```

> convergence(allSurroRes)
      maxSgrad minHev minREev
Clayton unadj  FALSE  FALSE   ---
Clayton adj   FALSE  FALSE   TRUE
5 Plackett unadj  FALSE  FALSE   ---
Plackett adj   FALSE  FALSE   TRUE
Hougaard unadj  FALSE  TRUE    ---
Hougaard adj   FALSE  TRUE    TRUE
PoissonT       TRUE  TRUE    FALSE
10 PoissonI     TRUE  TRUE    ---
PoissonTI      TRUE  TRUE    TRUE
PoissonTIIa    TRUE  TRUE    TRUE

```

If the values of the minimum gradient and of the maximum eigenvalues are needed, the function `convals()` can be used:

```

> convals(allSurroRes)
      maxSgrad  minHev  minREev
Clayton unadj  1.5e+00 -6.1e+00   ---
Clayton adj   1.5e+00 -6.1e+00  1.0e-02
5 Plackett unadj  2.3e+02 -5.2e+00   ---
Plackett adj   2.3e+02 -5.2e+00  8.9e-03
Hougaard unadj  1.4e+01  7.8e-01   ---
Hougaard adj   1.4e+01  7.8e-01  8.0e-03
PoissonT       1.3e-05  1.3e+02  6.3e-12

```

```

10 PoissonI      2.0e-05  6.8e+01   ---
PoissonTI      7.1e-06  6.7e+01  2.0e-02
PoissonTIIa    5.0e-05  9.4e+07  1.0e-01

```

### 3.2. Prediction of the treatment effect

When fitting surrogacy models, an estimate of the treatment effects on the two endpoints is computed for each trial. The function `predict()`, applied to an object of class `surrosurv`, returns the predictions of the treatment effects for each trial. The minimal syntax is `predict(allSurroRes)`, but one can be interested in prediction of only one of the fitted models:

```

> predict(allSurroRes, models = 'PoissonTI')
Treatment effect prediction for surrosurv
      object

      Poisson TI
      Trial      1      2      3      4

      Treatment effects on S:
      -0.52 -0.42 -0.38 -0.08 ...

10      Treatment effects on T:
      -0.26 -0.08 -0.27  0.41...

```

This function returns an object of class `predictSurrosurv`.

The predicted treatment effects can also be visualized graphically using the linear regression of the effect on  $T$  given the effect on  $S$ , provided that usual assumptions of linear regression hold. The usual surrogacy plot is obtained using the function `plot()` for the classes `surrosurv` and `predictSurrosurv`. For example, the surrogacy plots for the adjusted Clayton copula and the Poisson TI models in the advanced GASTRIC meta-analysis (Fig. 2) can be obtained as follows:

```

> plot(allSurroRes,
       c('Clayton adj', 'PoissonTI'))

```

The argument `surro.stats` controls whether the estimated Kendall’s  $\tau$  and  $R^2$  must be displayed on the plots; `pred.ints` controls whether the prediction intervals must be plotted; `show.ste` controls whether the surrogate threshold effect (STE) must be displayed on the plots. The STE is the minimal treatment effect to be observed on the surrogate endpoint  $S$  to predict a statistically significant effect on the true endpoint  $T$  at a given significance level, usually 0.05 [30]. The value of the STE estimated by each surrogacy model can be obtained via the function `ste()`, both in terms of regression parameter (beta) and in terms of hazard ratio (HR):

```

> ste(allSurroRes)
      beta  HR

```

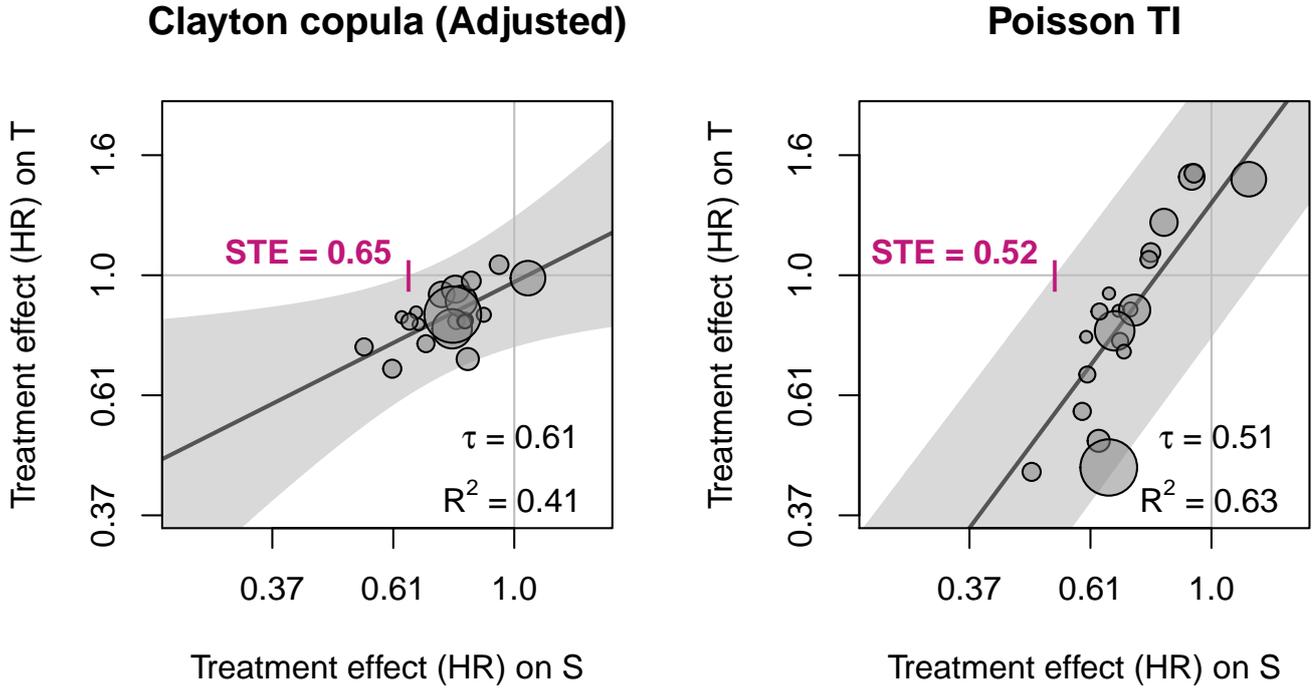


Figure 2: Predictions for the advanced GASTRIC meta-analysis, as computed by the adjusted Clayton copula model, which had poor convergence metrics, and by the Poisson TI model, which was deemed to have converged. HR = hazard ratio. STE = surrogate threshold effect.

```

Clayton.unadj -0.61 0.54
Clayton.adj -0.44 0.65
5 Plackett.unadj -0.61 0.54
Plackett.adj -4.15 0.02
Hougaard.unadj -0.61 0.54
Hougaard.adj -1.30 0.27
PoissonT -0.17 0.84
10 PoissonTI -0.65 0.52
PoissonTIa -1.16 0.31

```

imum between  $N$  and the maximum number of cores on the machine.

```

> loocvRes <- loocv(
  gastadv,
  models = c('Clayton',
             'PoissonTI'))
5 Parallel computing on 8 cores
  (the total number of cores detected)

```

### 3.2.1. Leave-one-trial-out cross-validation

One technique used to assess the validity of the surrogacy model is to apply the leave-one-out principle to the trials in the meta-analysis. This means that, for each trial, the observed treatment effect on  $S$  is compared to its prediction obtained by entering the observed effect on  $T$  in the surrogacy model fitted on the other  $N - 1$  trials. [31, 32, 33]. The function `loocv()` allows performing this evaluation for a given list of models. The cross-validation requires fitting as many models as the number of trials  $N$ . As each model is usually very time-consuming to converge, the function `loocv()` has been implemented to fit the  $N$  models by parallel computing. The argument `parallel` is a logical for allowing or not such a parallelization, whereas `nCores` allows specifying the number of cores to use. By default, `parallel = TRUE` and `nCores` is set to the min-

The results of the crossvalidation can be printed

```

> loocvRes
  Clayton copula (Unadjusted)
      1      2      3      4
4 obsBeta -0.31 -0.21 -0.09 -0.02 ...
  predict -0.40 -0.31 -0.07 -0.17 ...
  lwr     -0.76 -0.65 -0.42 -0.51 ...
  upr     -0.05  0.02  0.28  0.17 ...
  kTau    0.60  0.60  0.61  0.60 ...
9 R2      0.49  0.49  0.45  0.46 ...

  Clayton copula (Adjusted)
      1      2      3      4
14 obsBeta -0.31 -0.21 -0.09 -0.02 ...
  predict -0.39 -0.31 -0.09 -0.18 ...
  lwr     -0.69 -0.57 -0.35 -0.41 ...
  upr     -0.09 -0.04  0.17  0.06 ...
  kTau    0.60  0.60  0.61  0.60 ...

```

```

R2          0.46  0.45  0.42  0.46  ...
19
  Poisson TI
      1      2      3      4
obsBeta -0.31 -0.21 -0.09 -0.02 ...
predict -0.69 -0.40  0.08 -0.11 ...
24 lwr    -1.22 -1.00 -0.76 -0.68 ...
   upr    -0.15  0.20  0.92  0.47 ...
kTau     0.51  0.52  0.51  0.51 ...
R2       0.70  0.65  0.47  0.60 ...

```

and plotted (Fig. 3) by showing, for each trial, the comparison between the observed treatment effect on  $T$ , and its prediction interval, based on the observed treatment effect on  $S$  for the same trial and the surrogacy model fitted on the other  $N - 1$  trials:

```
> plot(loocvRes)
```

### 3.3. Utilities for data simulation

Few publications present simulation approaches adapted to discuss statistical methods for evaluating failure time surrogate endpoints [6, 34, 35, 36, 37]. To our knowledge, the data generation methods used to date are based either on the use of a Clayton copula or on a mixture of half-normal and exponential random variables. Thanks to the `surroSurv` package, data can be generated using these two methods, in addition to an approach based on mixed proportional hazard models that we employed recently [9]. These three data generation algorithms are detailed here below. Of note, simulated data are useful to check the operating characteristics of new statistical methods for surrogate end point validation. Conversely, simulated data do not play any role for validating surrogate endpoints in medical applications.

#### 3.3.1. Data generation based on a Clayton copula

The data generation method used in [6] and in [36, 37] reflects the data generating process underlying the two-step copula model (Sec. 2.1).

We implemented this approach for the Clayton family (Eq. (2)), which is available using the function `simData.cc()`. This function generates data as follows:

- trial-specific random effects are generated from

$$\begin{pmatrix} m_{S_i} \\ m_{T_i} \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_S^2 & \sigma_S \sigma_T \rho_m \\ \sigma_S \sigma_T \rho_m & \sigma_T^2 \end{pmatrix} \right)$$

- trial-specific treatment effects are generated from

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} d_a^2 & d_a d_b \rho_{\text{trial}} \\ d_a d_b \rho_{\text{trial}} & d_b^2 \end{pmatrix} \right)$$

- exponentially distributed individual times are simulated for  $S$ , conditionally on the random effects generated before.

$$S_{ij} = -\log(U_{S_{ij}})/\lambda_{S_{ij}},$$

with

$$\lambda_{S_{ij}} = \exp(\mu_S + m_{S_i} + \alpha_i Z_{ij}),$$

$$U_{S_{ij}} \sim U(0, 1)$$

- exponentially distributed individual times are simulated for  $T | S$ , conditionally on the random effects generated before *and on the value of  $S$*

$$T_{ij} | S_{ij} = -\log(U'_{T_{ij}})/\lambda_{T_{ij}},$$

with

$$\lambda_{T_{ij}} = \exp(\mu_T + m_{T_i} + \beta_i Z_{ij}),$$

$$U'_{T_{ij}} = \left[ \left( U_{T_{ij}}^{-\theta/(1+\theta)} - 1 \right) U_{S_{ij}}^{-\theta} + 1 \right]^{-1/\theta},$$

$$U_{T_{ij}} \sim U(0, 1).$$

The details of the arguments of the `simData.cc()` function can be obtained using `help(simData.cc)`.

#### 3.3.2. Data generation based on a mixture of half-normal and exponential random variables

The data generation method used in [34] and in [35] is based on the results by Cowles [38], which showed that a Weibull distribution can be expressed as a scaled mixture of half-normal distribution and an exponential distribution with unit rate parameter.

This approach is implemented in the function `simData.mx()` and generates data as follows:

- trial-specific random effects are generated from

$$\begin{pmatrix} m_{S_i} \\ m_{T_i} \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_S^2 & \sigma_S \sigma_T \rho_m \\ \sigma_S \sigma_T \rho_m & \sigma_T^2 \end{pmatrix} \right)$$

- trial-specific treatment effects are generated from

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} d_a^2 & d_a d_b \rho_{\text{trial}} \\ d_a d_b \rho_{\text{trial}} & d_b^2 \end{pmatrix} \right)$$

- individual half-normal random variables  $Y_{ij}^*$  are generated from the distribution

$$f(y^*) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{y^{*2}}{2}\right), \quad y^* \in \mathbb{R}_+$$

- unit rate parameter exponential random variables  $\Lambda_{S_{ij}}$  and  $\Lambda_{T_{ij}}$  are generated from  $-\log(U_{S_{ij}})_{S_{ij}}$  and  $-\log(U_{T_{ij}})$ , with  $U_{S_{ij}} \sim U(0, 1)$  and  $U_{T_{ij}} \sim U(0, 1)$

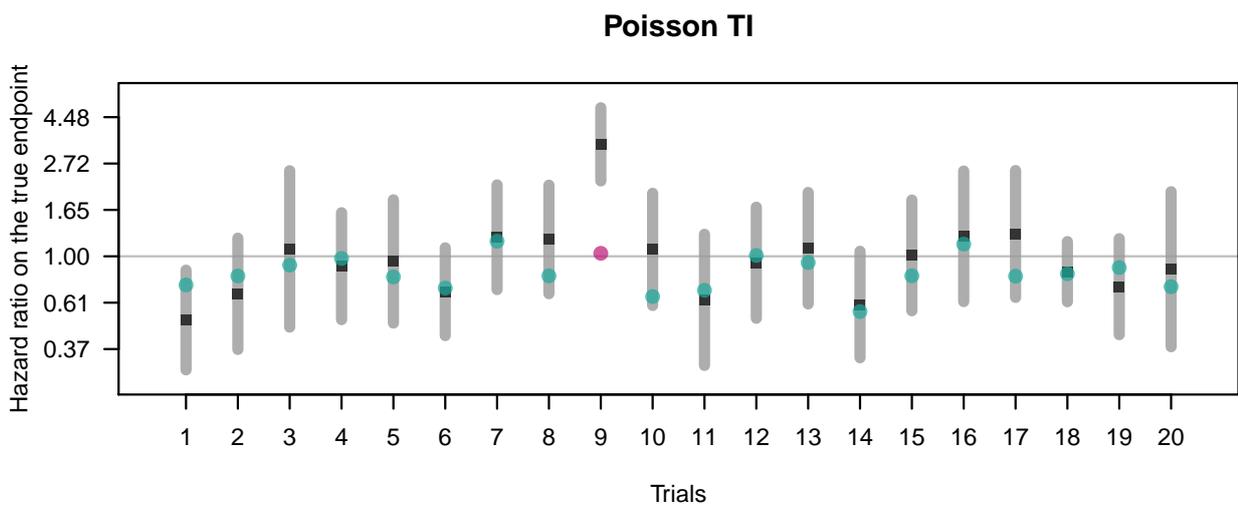
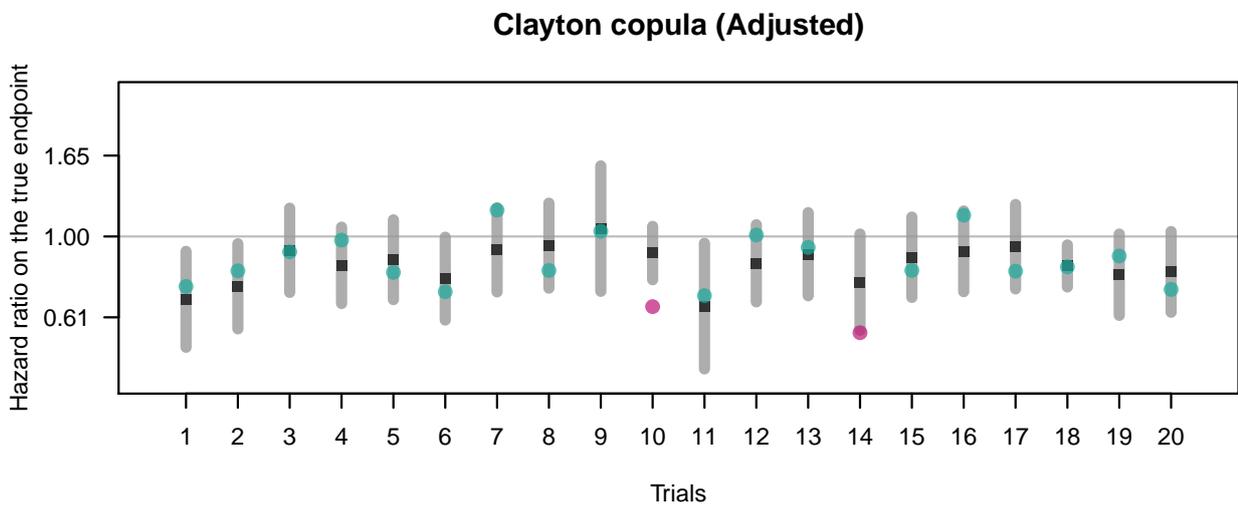
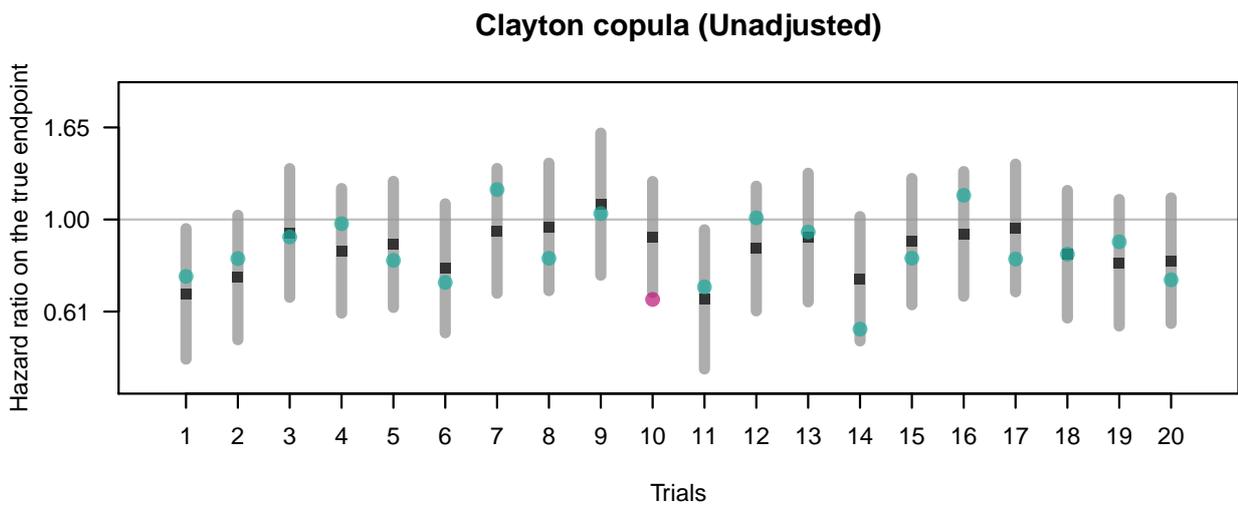


Figure 3: Leave-one-trial-out cross-validation results for the advanced GASTRIC meta-analysis. The black squares and the vertical grey lines are the predicted values of the treatment effect on overall survival (OS), with the 95% prediction intervals (PI). Dots are the observed treatment effects on OS (green = within the PI, magenta = out of the PI).

- exponentially distributed individual times are simulated for  $S$  and  $T$  from

$$S_{ij} = \left( Y_{ij}^* \sqrt{2\Lambda_{Sij}} \right) \exp(\mu_S + m_{S_i} + \alpha_i Z_{ij}),$$

$$T_{ij} = \left( Y_{ij}^* \sqrt{2\Lambda_{Tij}} \right) \exp(\mu_S + m_{T_i} + \alpha_i Z_{ij}).$$

The details of the arguments can be obtained using `help(simData.mx)`.

### 3.3.3. Data generation based on mixed proportional hazard models

Recently we also generated data using individual random effects to control individual-level surrogacy [9]. This approach is implemented in the function `simData.re()` and generates data as follows:

- trial-specific random effects and trial-specific treatment effects were generated as in the Clayton copula case
- individual random effects were generated from  $u_{ij} \sim \mathcal{N}(0, \sigma^2)$ , with  $\sigma^2$  depending on the scenario (according to the Kendall's  $\tau$ )
- exponentially distributed individual times were simulated for  $S$  and  $T$ , conditionally on the random effects generated before. We used the inverse transform method, which consists in transforming a uniform random variable by means of the inverse of the probability distribution function of the random variable to be generated [see for instance 39, § 2.1.2]

$$S_{ij} = -\log(U_{Sij})/\lambda_{Sij},$$

with

$$\lambda_{Sij} = \exp(\mu_S + m_{S_i} + \alpha_i Z_{ij} + u_{ij}),$$

$$U_{Sij} \sim U(0, 1),$$

and

$$T_{ij} = -\log(U_{Tij})/\lambda_{Tij},$$

with

$$\lambda_{Tij} = \exp(\mu_T + m_{T_i} + \beta_i Z_{ij} + u_{ij}),$$

$$U_{Tij} \sim U(0, 1).$$

The details of the arguments can be obtained using `help(simData.re)`.

## 4. Mode of availability of the `surrosurv` package

The `surrosurv` package is an open-source project. Stable versions are released via the Comprehensive R Archive Network (CRAN, <https://cran.r-project.org/package=surrosurv>). Source code is available on GitHub (<https://github.com/Oncostat/surrosurv>).

## 5. Discussion

An increasing interest can be observed in surrogate endpoints, mostly as they can accelerate drawing valid conclusion from clinical trials, thus reducing development costs and provide patients with new effective treatments more quickly [40]. Composite endpoints, i. e. including the true endpoint in addition to further events, have been challenged because of their interpretability [41] or their likeliness to overestimate the effect of new treatments [42]. Despite, they remain largely used as surrogate endpoints in medical fields such as oncology and cardiovascular diseases because of their meaningfulness for clinicians and because they are expected to better correlate with the final outcome while increasing statistical power as compared to this latter [43, 44, 45].

The validation of surrogate endpoints for clinical trials relies on the principles that the surrogate must correlate with the true endpoint and that the treatment effect on the surrogate must correlate with the effect on the true endpoint [46, 47]. Consensus has been reached about the importance of using data from several randomized controlled trials and to combine them in a meta-analytic framework [46, 2, 48, 1, 27]. Despite the usefulness of bringing several trials of similar treatments together in the analysis, an excess of heterogeneity of the study populations, of the experimental and control treatments could affect the interpretability of the results. Conversely, there needs to be a range of treatment effects for the meta-analytic approach to be useful and to reliably extrapolate the results to new trials.

However, different statistical measures can be used for assessing the validity of surrogate endpoints and there remains uncertainty about the quantitative definition of a valid surrogate marker [49, 50]. The case of failure time endpoints is particularly challenging from a statistical and computational viewpoint [51, 52]. Two-step copula models [3] are the state-of-the-art statistical methodology, commonly used in numerous clinical publications over the last 15 years [53, 54, 55, 56, 31, 5, 14, 32, 57, 33, 58, 59], including a surrogate evaluation conducted by the US FDA [60]. Recently, we also proposed an alternative estimation strategy [9], based on a single step estimation as in the gaussian case [2]. This alternative approach exploits the equivalence of parameter estimates between proportional hazard models and auxiliary Poisson log-linear models.

Statistical software implementing the two-step copula approach is available in SAS macros [61], allowing to estimate unadjusted and adjusted  $R^2$  as well as individual level correlation for Clayton, Plackett, and Hougaard copulas. The R package `Surrogate` [62, 63] implements the information-theoretic approach [64, 65], with a weighted linear regression at the second step. The `surrosurv` package is the first to provide full implementation of the classical approach by Burzykowski et al. [3], with and without adjustment for estimation error in the second step. In addition, our package implements the recent Poisson

approach presented in [9]. As compared to previously available software, the `surrosurv` package also provides utilities for prediction of treatment effect for new trials, plots with prediction intervals, leave-one-trial-out cross-validation, data simulation, and convergence diagnostics.

Two convergence metrics are provided, based on the Karush–Kuhn–Tucker conditions [66]: maximum gradient and minimum eigenvalue of the covariance matrix. The first criterion requires that the (maximum) slope of the negative likelihood at the estimate is close to zero, i. e. the point is a (at least local) minimum. The second condition checks that the (minimum) eigenvalue of the Hessian matrix at the estimate is positive and far from zero, i. e. the Hessian is positive definite and the variances are strictly positive. This ensures that the maximum point is really a minimum of the negative likelihood and not a saddle point. Thresholds for a small-enough gradient and big-enough eigenvalue are strictly arbitrary and no guidance exists, to the best of our knowledge. Based on our personal experience, we set the default values to  $1e-2$  and  $1e-8$ , respectively, but the users can choose their own thresholds. Furthermore, the function `convvals()` returns the values of the two criteria, without comparison to any prespecified threshold. For the Poisson approach, the package also provides reduced models, which can be used in the case of an application for which the complete model seems overparametrized or too few information is available to converge to a reliable estimate of all the parameters.

In this paper we present the `surrosurv` package which provides an R implementation of classical and recent statistical methods for surrogacy assessment of failure time endpoints. In addition to implementation of different methods, the `surrosurv` package provides many additional features as compared to presently available software: utilities are available to fit surrogacy models, assess their convergence for reliable results, estimate surrogacy measures at individual and trial levels, estimate the surrogate threshold effect (STE), and perform leave-one-trial-out cross-validation. In addition, flexible simulation functions are can be used to generate data according to different methods described in the literature.

## Funding

The present work has been supported by the Institut National du Cancer (INCa), Grant SHS 2014-141, and by the Ligue Nationale Contre le Cancer. The study sponsors had no involvement in either the study design; the collection, analysis and interpretation of data; the writing of the manuscript; nor in the decision to submit the manuscript for publication.

## Acknowledgments

The authors thank the GASTRIC (Global Advanced/Adjuvant Stomach Tumor Research International Collabora-

tion) Group for permission to use their data. The investigators who contributed to GASTRIC are listed in references [14, 13]. The GASTRIC Group data are available within the `surrosurv` package for research purposes, under the conditions that (1) the research be scientifically appropriate, (2) the confidentiality of individual patient data be protected, (3) the results of the analyses be shared with the GASTRIC Group prior to public communication, (4) the source of data be fully acknowledged as above, and (5) resulting data and results be further shared with the research community.

- [1] T. Burzykowski, G. Molenberghs, M. Buyse, *The Evaluation of Surrogate Endpoints*, Springer Science & Business Media, New York, NY, 2006.
- [2] M. Buyse, G. Molenberghs, T. Burzykowski, D. Renard, H. Geys, The validation of surrogate endpoints in meta-analyses of randomized experiments, *Biostatistics* 1 (1) (2000) 49–67. doi:10.1093/biostatistics/1.1.49.
- [3] T. Burzykowski, G. Molenberghs, M. Buyse, H. Geys, D. Renard, Validation of surrogate end points in multiple randomized clinical trials with failure time end points, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 50 (4) (2001) 405–422. doi:10.1111/1467-9876.00244.
- [4] M. G. Kendall, A new measure of rank correlation, *Biometrika* 30 (1/2) (1938) 81–93. URL <http://www.jstor.org/stable/2332226>
- [5] K. Oba, X. Paoletti, S. Alberts, Y.-J. Bang, J. Benedetti, H. Bleiberg, P. Catalano, F. Lordick, S. Michiels, S. Morita, Y. Ohashi, J.-P. Pignon, P. Rougier, M. Sasako, J. Sakamoto, D. Sargent, K. Shitara, E. Van Cutsem, M. Buyse, T. Burzykowski, Disease-free survival as a surrogate for overall survival in adjuvant trials of gastric cancer: A meta-analysis, *Journal of the National Cancer Institute* 105 (21) (2013) 1600–1607. doi:10.1093/jnci/djt270.
- [6] T. Burzykowski, J. Cortiñas Abrahantes, Validation in the case of two failure-time endpoints, in: T. Burzykowski, G. Molenberghs, M. Buyse (Eds.), *The Evaluation of Surrogate Endpoints*, Springer, New York, NY, 2005, pp. 163–194.
- [7] J. Whitehead, Fitting Cox’s regression model to survival data using GLIM, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 23 (3) (1980) 268–275. doi:10.2307/2346901.
- [8] N. Laird, D. Olivier, Covariance analysis of censored survival data using log-linear analysis techniques, *Journal of the American Statistical Association* 76 (374) (1981) 231–40. doi:10.1080/01621459.1981.10477634.
- [9] F. Rotolo, X. Paoletti, M. Buyse, T. Burzykowski, S. Michiels, A Poisson approach to the validation of failure time surrogate endpoints in individual patient data meta-analyses, *Statistical Methods in Medical Research* Epub ahead of press. doi:10.1177/0962280217718582.
- [10] L. Duchateau, P. Janssen, *The Frailty Model*, Springer Verlag, New York, NY, 2008. doi:10.1007/978-0-387-72835-3.
- [11] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2016). URL <https://www.R-project.org/>
- [12] F. Rotolo, `surrosurv`: Evaluation of failure time surrogate endpoints in individual patient data meta-analyses, *R package version 1.1.24* (2017). URL <https://CRAN.R-project.org/package=surrosurv>
- [13] GASTRIC group, Role of chemotherapy for advanced/recurrent gastric cancer: An individual-patient-data meta-analysis, *European Journal of Cancer* 49 (7) (2013) 1565–1577. doi:10.1016/j.ejca.2012.12.016.
- [14] X. Paoletti, K. Oba, Y.-J. Bang, H. Bleiberg, N. Boku, O. Bouché, P. Catalano, N. Fuse, S. Michiels, M. Moehler, S. Morita, Y. Ohashi, A. Ohtsu, A. Roth, P. Rougier,

- J. Sakamoto, D. Sargent, M. Sasako, K. Shitara, P. Thuss-Patience, E. Van Cutsem, T. Burzykowski, M. Buyse, Progression-free survival as a surrogate for overall survival in advanced/recurrent gastric cancer trials: A meta-analysis, *Journal of the National Cancer Institute* 105 (21) (2013) 1667–1670. doi:10.1093/jnci/djt269.
- [15] M. Hofert, I. Kojadinovic, M. Maechler, J. Yan, **copula**: Multivariate dependence with copulas, R package version 0.999-15 (2016).  
URL <https://CRAN.R-project.org/package=copula>
- [16] J. Yan, Enjoy the joy of copulas: With a package **copula**, *Journal of Statistical Software* 21 (1) (2007) 1–21. doi:10.18637/jss.v021.i04.
- [17] D. Clayton, A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika* 65 (1) (1978) 141–151. doi:10.1093/biomet/65.1.141.
- [18] R. L. Plackett, A class of bivariate distributions, *Journal of the American Statistical Association* 60 (310) (1965) 516–522. doi:10.1080/01621459.1965.10480807.
- [19] P. Hougaard, A class of multivariate failure time distributions, *Biometrika* 73 (1986) 671–678. doi:10.1093/biomet/73.3.671.
- [20] H. C. van Houwelingen, L. Arends, T. Stijnen, Advanced methods in meta-analysis: Multivariate approach and meta-regression, *Statistics in Medicine* 21 (2002) 589–624. doi:10.1002/sim.1040.
- [21] A. Gasparrini, B. Armstrong, M. Kenward, Multivariate meta-analysis for non-linear and other multi-parameter associations, *Statistics in Medicine* 31 (29) (2012) 3821–3839. doi:10.1002/sim.5471.
- [22] A. Gasparrini, **mvmeta**: Multivariate and univariate meta-analysis and meta-regression, R package version 0.4-7 (2015).  
URL <https://CRAN.R-project.org/package=mvmeta>
- [23] M. J. Crowther, R. D. Riley, J. A. Staessen, J. Wang, F. Gueyffier, P. C. Lambert, Individual patient data meta-analysis of survival data using poisson regression models, *BMC Medical Research Methodology* 12 (1) (2012) 34. doi:10.1186/1471-2288-12-34.
- [24] C. Goutis, G. Casella, Explaining the saddlepoint approximation, *The American Statistician* 53 (3) (1999) 216–224. doi:10.1080/00031305.1999.10474463.
- [25] M. Munda, F. Rotolo, C. Legrand, **parfm**: Parametric frailty models in R, *Journal of Statistical Software* 51 (1) (2012) 1–20. doi:10.18637/jss.v051.i11.
- [26] F. Rotolo, M. Munda, **parfm**: Parametric frailty models, R package version 2.7.5 (2017).  
URL <https://CRAN.R-project.org/package=parfm>
- [27] M. Buyse, G. Molenberghs, X. Paoletti, K. Oba, A. Alonso, W. Van der Elst, T. Burzykowski, Statistical evaluation of surrogate endpoints with examples from cancer clinical trials, *Biometrical Journal* 58 (1) (2016) 104–132. doi:10.1002/bimj.201400049.
- [28] J. C. Nash, On best practice optimization methods in R, *Journal of Statistical Software* 60 (2) (2012) 1–14. doi:10.18637/jss.v043.i09.
- [29] J. C. Nash, R. Varadhan, G. Grothendieck, **optimx**: A replacement and extension of the **optim()** function, R package version 2013.8.7 (2013).  
URL <https://CRAN.R-project.org/package=optimx>
- [30] T. Burzykowski, M. Buyse, Surrogate threshold effect: An alternative measure for meta-analytic surrogate endpoint validation, *Pharmaceutical Statistics* 5 (3) (2006) 173–186. doi:10.1002/pst.207.
- [31] S. Michiels, A. Le Maître, M. Buyse, T. Burzykowski, E. Maillard, J. Bogaerts, J. B. Vermorken, W. Budach, T. F. Pajak, K. K. Ang, J. Bourhis, J.-P. Pignon, Surrogate endpoints for overall survival in locally advanced head and neck cancer: meta-analyses of individual patient data, *The Lancet Oncology* 10 (4) (2009) 341–50. doi:10.1016/S1470-2045(09)70023-3.
- [32] A. Mauguen, J.-P. Pignon, S. Burdett, C. Domerg, D. Fisher, R. Paulus, S. J. Mandrekar, C. P. Belani, F. A. Shepherd, T. Eisen, H. Pang, L. Collette, W. T. Sause, S. E. Dahlberg, J. Crawford, M. O’Brien, S. E. Schild, M. Parmar, J. F. Tierney, C. Le Pechoux, S. Michiels, Surrogate endpoints for overall survival in chemotherapy and radiotherapy trials in operable and locally advanced lung cancer: A re-analysis of meta-analyses of individual patients’ data, *The Lancet Oncology* 14 (7) (2013) 619–26. doi:10.1016/S1470-2045(13)70158-X.
- [33] F. Rotolo, J.-P. Pignon, J. Bourhis, S. Marguet, J. Leclercq, W. T. Ng, J. Ma, A. Chan, P.-Y. Huang, G. Zhu, D. Chua, Y. Chen, H. Q. Mai, D. L. W. Kwong, Y. L. Soong, J. Moon, Y. Tung, K. H. Chi, G. Fountzilas, L. Zhang, E. Pun Hui, A. W. M. Lee, P. Blanchard, S. Michiels, Surrogate endpoints for overall survival in loco-regionally advanced nasopharyngeal carcinoma: an individual patient data meta-analysis, *Journal of the National Cancer Institute* 109 (4). doi:10.1093/jnci/djw239.
- [34] Q. Shi, L. A. Renfro, B. M. Bot, T. Burzykowski, M. Buyse, D. J. Sargent, Comparative assessment of trial-level surrogacy measures for candidate time-to-event surrogate endpoints in clinical trials, *Computational Statistics & Data Analysis* 55 (9) (2011) 2748–2757. doi:10.1016/j.csda.2011.03.014.
- [35] L. A. Renfro, Q. Shi, D. J. Sargent, B. P. Carlin, Bayesian adjusted  $R^2$  for the meta-analytic evaluation of surrogate time-to-event endpoints in clinical trials, *Statistics in Medicine* 31 (8) (2012) 743–761. doi:10.1002/sim.4416.
- [36] L. A. Renfro, Q. Shi, Y. Xue, J. Li, H. Shang, D. J. Sargent, Center-within-trial versus trial-level evaluation of surrogate endpoints, *Computational Statistics & Data Analysis* 78 (Supplement C) (2014) 1–20. doi:10.1016/j.csda.2014.03.011.
- [37] L. A. Renfro, H. Shang, D. J. Sargent, Impact of copula directional specification on multi-trial evaluation of surrogate endpoints, *Journal of Biopharmaceutical Statistics* 25 (4) (2015) 857–877. doi:10.1080/10543406.2014.920870.
- [38] M. Cowles, Evaluating surrogate endpoints for clinical trials: A bayesian approach, Tech. rep., University of Iowa (2004).  
URL <https://stat.uiowa.edu/sites/stat.uiowa.edu/files/techrep/tr334.pdf>
- [39] C. Robert, G. Casella, *Introducing Monte Carlo Methods with R*, Springer Science & Business Media, 2009. doi:10.1007/978-1-4419-1576-4.
- [40] T. R. Fleming, Surrogate endpoints and FDA’s accelerated approval process, *Health Affairs* 24 (1) (2005) 67–78. doi:10.1377/hlthaff.24.1.67.
- [41] T. R. Fleming, J. H. Powers, Biomarkers and surrogate endpoints in clinical trials, *Statistics in Medicine* 31 (25) (2012) 2973–2984. doi:10.1002/sim.5403.
- [42] O. Ciani, M. Buyse, R. Garside, T. Pavey, K. Stein, J. A. Sterne, R. S. Taylor, Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study, *The BMJ* 346 (2013) f457. doi:10.1136/bmj.f457.
- [43] C. A. Bellera, M. Pulido, S. Gourgou, L. Collette, A. Dous-sau, A. Kramar, T. S. Dabakuyo, M. Ouali, A. Auperin, T. Filleron, C. Fortpied, C. Le Tourneau, X. Paoletti, M. Mauer, S. Mathoulin-Plissier, F. Bonnetain, Protocol of the Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATECAN) project: formal consensus method for the development of guidelines for standardised time-to-event endpoints’ definitions in cancer clinical trials, *European Journal of Cancer* 49 (4) (2013) 769–781. doi:10.1016/j.ejca.2012.09.035.
- [44] F. Bonnetain, B. Bonsing, T. Conroy, A. Dousseau, B. Glimelius, K. Haustermans, F. Lacaine, J. L. V. Laethem, T. Aparicio, D. Aust, C. Bassi, V. Berger, E. Chamorey, B. Chibaudel, L. Dahan, A. D. Gramont, J. R. Delpero, C. Dervenis, M. Ducreux, J. Gal, E. Gerber, P. Ghaneh, P. Hammel, A. Hendlisz, V. Jooste, R. Labianca, A. Latouche, M. Lutz, T. Macarulla, D. Malka, M. Mauer, E. Mitry, J. Neoptolemos, P. Pessaux, A. Sauvanet, J. Taberero, J. Taieb, G. van Tienhoven, S. Gourgou-Bourgade, C. Bellera, S. Mathoulin-Plissier, L. Collette, Guidelines for time-to-event end-point definitions in trials for pancreatic cancer. results of the DATECAN initia-

- tive (Definition for the Assessment of Time-to-event End-points in CANcer trials), *European Journal of Cancer* 50 (17) (2014) 2983–2993. doi:<https://doi.org/10.1016/j.ejca.2014.07.011>.
- [45] S. Gourgu-Bourgade, D. Cameron, P. Poortmans, B. Asselain, D. Azria, F. Cardoso, R. A'Hern, J. Bliss, J. Bogaerts, H. Bonnefoi, E. Brain, M. J. Cardoso, B. Chibaudel, R. Coleman, T. Cufer, L. Dal Lago, F. Dalenc, E. De Azambuja, M. Debled, S. Delaloge, T. Filleron, J. Gligorov, M. Gutowski, W. Jacot, C. Kirkove, G. MacGrogan, S. Michiels, I. Negreiros, B. V. Offerens, F. Penault Llorca, G. Pruneri, H. Roche, N. S. Russell, F. Schmitt, V. Servent, B. Thrlimann, M. Untch, J. A. van der Hage, G. van Tienhoven, H. Wildiers, J. Yarnold, F. Bonnetain, S. Mathoulin-Pliessier, C. Bellera, T. S. Dabakuyo-Yonli, Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials), *Annals of Oncology* 26 (5) (2015) 873–879. doi:[10.1093/annonc/mdv106](https://doi.org/10.1093/annonc/mdv106).
- [46] M. J. Daniels, M. D. Hughes, Meta-analysis for the evaluation of potential surrogate markers, *Statistics in Medicine* 16 (17) (1997) 1965–1982. doi:[10.1002/\(SICI\)1097-0258\(19970915\)16:17<1965::AID-SIM630>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1097-0258(19970915)16:17<1965::AID-SIM630>3.0.CO;2-M).
- [47] M. Buyse, G. Molenberghs, Criteria for the validation of surrogate endpoints in randomized experiments, *Biometrics* 54 (3) (1998) 1014–29. doi:[10.2307/2533853](https://doi.org/10.2307/2533853).
- [48] M. H. Gail, R. Pfeiffer, H. C. Van Houwelingen, R. J. Carroll, On meta-analytic assessment of surrogate outcomes, *Biostatistics* 1 (3) (2000) 231–46. doi:[10.1093/biostatistics/1.3.231](https://doi.org/10.1093/biostatistics/1.3.231).
- [49] M. N. Lassere, K. R. Johnson, M. Schiff, D. Rees, Is blood pressure reduction a valid surrogate endpoint for stroke prevention? an analysis incorporating a systematic review of randomised controlled trials, a by-trial weighted errors-in-variables regression, the surrogate threshold effect (STE) and the biomarker-surrogacy (BioSurrogate) evaluation schema (BSES), *BMC Medical Research Methodology* 12 (27) (2012) 1–21. doi:[10.1186/1471-2288-12-27](https://doi.org/10.1186/1471-2288-12-27).
- [50] K. R. Johnson, N. Freemantle, D. M. Anthony, M. N. Lassere, LDL-cholesterol differences predicted survival benefit in statin trials by the surrogate threshold effect (STE), *Journal of Clinical Epidemiology* 62 (3) (2009) 328–336. doi:<https://doi.org/10.1016/j.jclinepi.2008.06.004>.
- [51] T. Burzykowski, Surrogate endpoints: wishful thinking or reality?, *Statistical Methods in Medical Research* 17 (5) (2008) 463–466. doi:[10.1177/0962280207081866](https://doi.org/10.1177/0962280207081866).
- [52] E. Green, G. Yothers, D. J. Sargent, Surrogate endpoint validation: statistical elegance versus clinical relevance, *Statistical Methods in Medical Research* 17 (5) (2008) 477–486. doi:[10.1177/0962280207081863](https://doi.org/10.1177/0962280207081863).
- [53] D. J. Sargent, H. S. Wieand, D. G. Haller, J. K. Benedetti, M. Buyse, R. Labianca, J. F. Seitz, C. J. O'Callaghan, G. Francini, A. Grothey, M. O'Connell, P. J. Catalano, C. D. Blanke, D. Kerr, E. Green, N. Wolmark, T. Andre, R. Goldberg, A. De Gramont, Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: Individual patient data from 20,898 patients on 18 randomized trials, *Journal of Clinical Oncology* 23 (34) (2005) 8664–8670. doi:[10.1200/JCO.2005.01.6071](https://doi.org/10.1200/JCO.2005.01.6071).
- [54] M. Buyse, T. Burzykowski, K. Carroll, S. Michiels, D. J. Sargent, L. L. Miller, G. L. Elfring, J.-P. Pignon, P. Piedbois, Progression-free survival is a surrogate for survival in advanced colorectal cancer, *Journal of Clinical Oncology* 25 (33) (2007) 5218–5224. doi:[10.1200/JCO.2007.11.8836](https://doi.org/10.1200/JCO.2007.11.8836).
- [55] T. Burzykowski, M. Buyse, M. J. Piccart-Gebhart, G. Sledge, J. Carmichael, H.-J. Lück, J. R. Mackey, J.-M. Nabholz, R. Paridaens, L. Biganzoli, J. Jassem, M. Bontenbal, J. Bonnetterre, S. Chan, G. A. Basaran, P. Therasse, Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer, *Journal of Clinical Oncology* 26 (12) (2008) 1987–1992. doi:[10.1200/JCO.2007.10.8407](https://doi.org/10.1200/JCO.2007.10.8407).
- [56] M. Buyse, T. Burzykowski, S. Michiels, K. Carroll, Individual- and trial-level surrogacy in colorectal cancer, *Statistical Methods in Medical Research* 17 (5) (2008) 467–75. doi:[10.1177/0962280207081864](https://doi.org/10.1177/0962280207081864).
- [57] S. Michiels, L. Pugliano, S. Marguet, D. Grun, J. Barinoff, D. Cameron, M. Cobleigh, A. Di Leo, S. Johnston, G. Gasparini, et al., Progression-free survival as surrogate end point for overall survival in clinical trials of HER2-targeted agents in HER2-positive metastatic breast cancer, *Annals of Oncology* 27 (6) (2016) 1029–1034. doi:[10.1093/annonc/mdw132](https://doi.org/10.1093/annonc/mdw132).
- [58] W. Xie, M. M. Regan, M. Buyse, S. Halabi, P. W. Kantoff, O. Sartor, H. Soule, N. W. Clarke, L. Collette, J. J. Dignam, et al., Metastasis-free survival is a strong surrogate of overall survival in localized prostate cancer, *Journal of Clinical Oncology* 35 (27) (2017) 3097–3104. doi:[10.1200/JCO.2017.73.9987](https://doi.org/10.1200/JCO.2017.73.9987).
- [59] S. Michiels, E. D. Saad, M. Buyse, Progression-free survival as a surrogate for overall survival in clinical trials of targeted therapy in advanced solid tumors, *Drugs* 77 (7) (2017) 713–719. doi:[10.1007/s40265-017-0728-y](https://doi.org/10.1007/s40265-017-0728-y).
- [60] G. M. Blumenthal, S. W. Karuri, H. Zhang, L. Zhang, S. Khozin, D. Kazandjian, S. Tang, R. Sridhara, P. Keegan, R. Pazdur, Overall response rate, progression-free survival, and overall survival with targeted and standard therapies in advanced non-small-cell lung cancer: US Food and Drug Administration trial-level and patient-level analyses, *Journal of Clinical Oncology* 33 (9) (2015) 1008–1014. doi:[10.1200/JCO.2014.59.0489](https://doi.org/10.1200/JCO.2014.59.0489).
- [61] T. Burzykowski, Two failure-time endpoints, in: A. Alonso, T. Bigirimurame, T. Burzykowski, M. Buyse, G. Molenberghs, L. Muchene, N. Joy Perualila, Z. Shkedy, W. Van der Elst (Eds.), *Applied Surrogate Endpoint Evaluation Methods with SAS and R*, CRC Press, Boca Raton, FL, 2017, pp. 67–80.
- [62] W. Van der Elst, P. Meyvisch, A. Alonso, H. M. Ensor, C. J. W. G. Molenberghs, *Surrogate: Evaluation of surrogate endpoints in clinical trials*, R package version 0.5 (2017). URL <https://CRAN.R-project.org/package=Surrogate>
- [63] W. Van der Elst, A. Alonso Abad, G. Molenberghs, The R package Surrogate, in: A. Alonso, T. Bigirimurame, T. Burzykowski, M. Buyse, G. Molenberghs, L. Muchene, N. Joy Perualila, Z. Shkedy, W. Van der Elst (Eds.), *Applied Surrogate Endpoint Evaluation Methods with SAS and R*, CRC Press, Boca Raton, FL, 2017, pp. 215–248.
- [64] A. Alonso, G. Molenberghs, Surrogate marker evaluation from an information theory perspective, *Biometrics* 63 (1) (2007) 180–186. doi:[10.1111/j.1541-0420.2006.00634.x](https://doi.org/10.1111/j.1541-0420.2006.00634.x).
- [65] A. Alonso, G. Molenberghs, Evaluating time to cancer recurrence as a surrogate marker for survival from an information theory perspective, *Statistical Methods in Medical Research* 17 (5) (2008) 497–504. doi:[10.1177/0962280207081851](https://doi.org/10.1177/0962280207081851).
- [66] H. W. Kuhn, *Nonlinear programming: a historical view*, in: *Traces and emergence of nonlinear programming*, Springer, 2014, pp. 393–414.