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

Abstract

Surrogate endpoints are endpoints which are attractive for use in clinical trials instead of well-established endpoints because of practical convenience. Two measures of surrogacy can be obtained in a meta-analytic context: the individual-level R^2_{indiv} and the trial-level R^2_{trial}. In the case of failure time endpoints, the usual two-step approach estimates a Kendall's τ at the individual level using a copula model, and then computes the R^2_{trial} via a weighted regression. Recently, we also developped an approach based on bivariate survival models with individual random effects to measure the τ and treatment-by-trial interactions to measure the R^2_{trial}. We used auxiliary mixed Poisson models to fit such a model.

The R package **surrosurv** implements the two-step method with Clayton, Plackett, and Hougaard copulas (with and without measurement-error adjustment), as well as the mixed Poisson approach. In this paper, we present the package and we show its use in practice on individual patient data from a meta-analysis of 4069 patients with advanced/recurrent gastric cancer from 20 trials of chemotherapy.

Keywords: surrogate endpoint, survival, time-to-event, randomized clinical trials, individual patient data meta-anaysis, copula, mixed proportional hazard models, Poisson regression, **surrosurv**, R.

1. Introduction

Surrogate endpoints are endpoints which can reliably used instead of well-established endpoints and which yield improved practical convenience in terms of cost, rapidity or ease of assessment, or reduced invasiveness (Burzykowski *et al.* 2006). A reliable surrogate endpoint must be strongly associatied with the true endpoints at the individual level and the effect of the treatment on the surrogate must be assciated with the effect on the true endpoint. In a meta-analytic context, the usual measure of individual level surrogacy for continuous endpoints is the R^2_{indiv} and the ususal measure of trial level surrogacy is the R^2_{trial} (Buyse *et al.* 2000).

In the case of failure time (survival) endpoints, the classical methods developped for normally-distributed endpoints cannot be used because of right censoring. Burzykowski *et al.* (2001) proposed a meta-analytic model for failure time endpoints that measures individual level surrogacy in terms of Kendall (1938)'s τ and trial level surrogacy in terms of R^2_{trial}. Despite the numerous applications of this approach in the medical literature, difficults exists to fit the

whole model directly and a two-step procedure is commonly employed. Such a procedure has some limitations preventing its applicability in a number of applications, including convergence issues which can make the interpretation of the results difficult (Oba *et al.* 2013; Burzykowski and Cortiñas Abrahantes 2005). In principle, the most natural framework for adapting the meta-analytic approach by Buyse *et al.* (2000) to the survival case is the use of bivariate mixed proportional hazard models, also known as frailty models (Duchateau and Janssen 2008). Since the estimation of frailty models with complex structures of random effects is computationally intensive and can fail to converge, we previously exploited (Rotolo *et al.* 201x) the well-known connection between the proportional hazard models and the Poisson log-linear models (Whitehead 1980; Laird and Olivier 1981), which has also been used more recently for individual patient data meta-analyses (Crowther *et al.* 2012).

In the present paper, we show how all these models can be fitted by use of the R (R Core Team 2016) package **surrosurv** (Rotolo 2016). We illustrate the use of the available functions using individual data of a meta-analysis of 20 randomized trials of chemotherapy, including 4069 patients with advanced/recurrent gastric cancer (GASTRIC group 2013; Paoletti *et al.* 2013).

2. The evaluation of failure time surrogate endpoints

Let T_{ij} and S_{ij} be the times to the true and the surrogate endpoints, respectively, for patient $j \in \{1, \ldots, n_i\}$ in trial $i \in \{1, \ldots, N\}$. 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.* (2001) for failure time endpoints consists in two steps, one for the individual and one for the trial level.

Individual-level. At the first the bivariate proportional hazard model is defined by means of the marginal hazard functions and the copula function accounting for their dependence:

$$\begin{cases} h_{Sij}(s; Z_{ij}) = h_{Si}(s) \exp\left\{\alpha_i Z_{ij}\right\} \\ h_{Tij}(t; Z_{ij}) = h_{Ti}(t) \exp\left\{\beta_i Z_{ij}\right\} \\ C_\theta(S_{Sij}(s), S_{Tij}(t)) \end{cases} \tag{1}$$

where $h_{Si}(s)$ and $h_{Ti}(s)$ are the trial-specific baseline hazards, α_i and β_i the treatment effects, and $S_{Sij}(s)$ and $S_{Tij}(t)$ the survival functions associated to of h_{Tij} and h_{Tij}. The dependence parameter θ is reparametrized into the individual-level Kendall's τ, according to the copula function thanks to the `tau()` function in the **copula** package (Hofert *et al.* 2016; Yan 2007).

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

- the Clayton (1978) copula

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

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

- the Plackett (1965) copula

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

  with $\theta > 0$ and Kendall's τ computed numerically as no analytical expression is available;

- the Hougaard (1986) copula

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

  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

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

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \boldsymbol{D} = \begin{pmatrix} d_a^2 & d_a d_b \rho_{\text{trial}} \\ d_a d_b \rho_{\text{trial}} & d_b^2 \end{pmatrix}\right), \tag{6}$$

$$\begin{pmatrix} \epsilon_{ai} \\ \epsilon_{bi} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Omega}_i = \begin{pmatrix} \omega_{ai}^2 & \omega_{ai}\omega_{bi}\rho_{\epsilon i} \\ \omega_{ai}\omega_{bi}\rho_{\epsilon i} & \omega_{bi}^2 \end{pmatrix}\right). \tag{7}$$

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

The trial-level surrogacy measure is $R^2_{\text{trial}} = \rho^2_{\text{trial}}$. In practice, the ρ_{trial} is computed via a linear regression of the β_i's over the α_i's adjusted by measurement error by fixing the $\boldsymbol{\Omega}_i$'s at their estimates from the first step (van Houwelingen *et al.* 2002) by using the **mvmeta** package (Gasparrini *et al.* 2012; Gasparrini 2015). 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^2_{trial}, obtained using a linear regression — equivalent to fixing all the elements of $\boldsymbol{\Omega}_i$ equal to 0 — by weigthing the observations $(\alpha_i, \beta_i)'$ by the trial size, in order to account somehow indirectly and approximately for measurement error.

2.2. One-step mixed Poisson approach

Let us assume that the bivariate proportional hazard model given by the first two lines of equation (1) holds conditionally on an individual random effect $u_{ij} \sim \mathcal{N}(0, \sigma^2_{\text{indiv}})$:

$$\begin{cases} h_{Sij}(s \mid u_{ij}) = h_{Si}(s) \exp\left\{u_{ij} + \alpha_i Z_{ij}\right\} \\ h_{Tij}(t \mid u_{ij}) = h_{Ti}(t) \exp\left\{u_{ij} + \beta_i Z_{ij}\right\}. \end{cases} \tag{8}$$

Note that this corresponds to a shared frailty model with bivariate clusters (Duchateau and Janssen 2008). The shared frailty term u_{ij} accounts for individual level dependence.

It is well-known (see for instance Whitehead 1980; Crowther *et al.* 2012) 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 $k = 1, \ldots, K$. 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 (8) can be estimated via a bivariate mixed Poisson model

$$\begin{cases} \log\left(\mu_{Sij}^{(k)}\right) = \mu_{Si}^{(k)} + u_{ij} + \alpha_i Z_{ij} + \log\left(y_{Sij}^{(k)}\right) \\ \log\left(\mu_{Tij}^{(k)}\right) = \mu_{Ti}^{(k)} + u_{ij} + \beta_i Z_{ij} + \log\left(y_{Tij}^{(k)}\right) \end{cases} \tag{9}$$

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 use of the a shared frailty u_{ij}, the estimated variance of which is $\hat{\sigma}^2_{\text{indiv}}$, can be used to compute an estimate of the Kendall's $\hat{\tau} = 4\int_0^\infty s\mathcal{L}(s)\mathcal{L}^{(2)}(s)\mathrm{d}s - 1$, 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 (Goutis and Casella 1999), implemented in the `fr.lognormal()` function in the **parfm** package (Munda *et al.* 2012; Rotolo and Munda 2015).

Trial-level surrogacy. In model (9), the same assumptions (6) as in the two-step copula model are done for the trial-specific treatment effects Thus, the correlation ρ_{trial} between the two treatment effects provides us with the coefficient of determination $R^2_{\text{trial}} = \rho^2_{\text{trial}}$, also referred to simply as R^2.

Reduced Poisson models. In order to deal with convergence issues, the **surrosurv** package computes four reduced versions of the full model (9).

- Model **Poisson T** has random trial-treatment interactions α_i and β_i, but does not incoroporate individual effects ($u_{ij} \equiv 0$) and it has 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^2_{trial}.
- 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 τ.
- 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 τ and R^2_{trial}.
- 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 $m_i \sim \mathcal{N}(0, \sigma^2_m)$.

3. A data example within the surrosurv package

We illustrate the use of the function in the **surrosurv** package by means of the individual patient data of the advanced GASTRIC meta-analysis (GASTRIC group 2013; Paoletti *et al.* 2013).

```
R> library('surrosurv')

Loading required package: optimx

R> packageVersion('surrosurv')

[1] '1.1.4'
```

The individual data of the 4069 patients, already made public by Buyse *et al.* (2016), are also available directly in R in the **surrosurv** package:

```
R> data('gastadv')
R> nrow(gastadv)

[1] 4069
```

The data set contains the following variables:

```
R> names(gastadv)

[1] "timeT"    "statusT"  "statusS"  "timeS"    "trialref" "trt"      "id"
```

where `timeT` and `timeS` are the (possibly censored) times for overall survival (`T`) and for progression-fre 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 survival curves for overall survival, the true endpoint T, and progression-free survival, the candiddate surrogate S.

4. 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 containing the trial identifier;
- `trt`, the treatment arm, coded as `-0.5` *vs.* `0.5`;
- `id`, a factor containing the patient id;

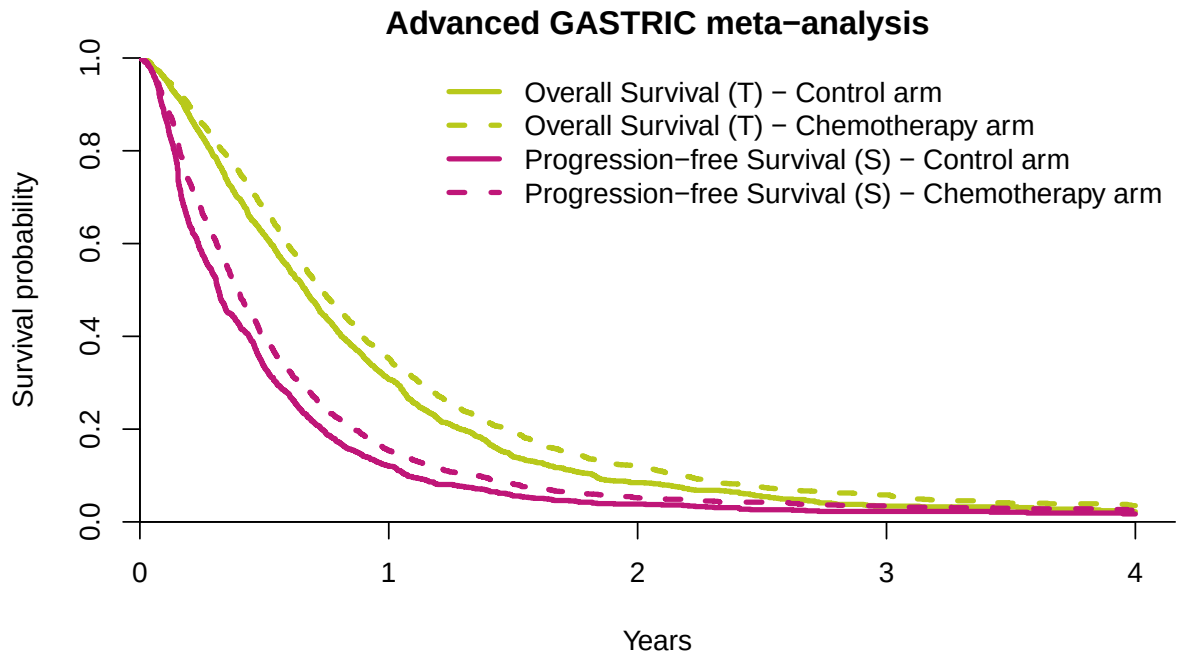

Figure 1: Survival curves for overall survival (T) and progression-free survival S in the advanced GASTRIC meta-analysis (GASTRIC group 2013)

- `timeT` and `timeS`, two positive-valued numerical variables, containing the observed or censoring 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. It can contain ny value of `clayton`, `plackett`, `hougaard`, or `poisson`). If not specified, all of them are fitted.

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

The optimizer used for optimization of the copula models and the Poisson models can be passed via the arguments `cop.OPTIMIZER` and `poi.OPTIMIZER`, passed to the **optimx** package (Nash 2012; Nash *et al.* 2013).

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:

```
R> allSurroRes <-  surrosurv(gastadv, verbose = TRUE)

Estimating model: clayton
```

```
Estimating model: plackett

Estimating model: hougaard

Estimating model: poisson
```

Note that, the computation time of the surrogacy models can be long. In this example, the computations required 38 mins on a PC with an Intel® 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 τ and R^2 can be easily shown:

```
R>   allSurroRes

               kTau R2
Clayton unadj  0.61 0.45
Clayton adj    0.61 0.42
Plackett unadj 0.62 0.45
Plackett adj   0.62 0.41
Hougaard unadj 0.32 0.45
Hougaard adj   0.32 0.38
PoissonT       -.-- 1
PoissonI       0.51 -.--
PoissonTI      0.51 0.63
PoissonTIa     0.51 0.83
```

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

4.1. Assessing convergence

The function `convergence()` checks whether convergence criteria are met by each of the fitted models:

```
R>   convergence(allSurroRes)

               maxSgrad minHev minREev
Clayton unadj     FALSE  FALSE     ---
Clayton adj       FALSE  FALSE    TRUE
Plackett unadj    FALSE  FALSE     ---
Plackett adj      FALSE  FALSE    TRUE
Hougaard unadj    FALSE   TRUE     ---
Hougaard adj      FALSE   TRUE    TRUE
PoissonT           TRUE   TRUE   FALSE
PoissonI           TRUE   TRUE     ---
PoissonTI          TRUE   TRUE    TRUE
PoissonTIa         TRUE   TRUE    TRUE
```

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 definitess of the two matrices. Two parameters can be used to tune the thresholds for 'small enough' maximum gradient and for 'big enough' minmum eigen value: `kkttol` (`1e-2` by default), and `kkt2tol` (`1e-8` by default).

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

```
R>   convals(allSurroRes)
```

```
               maxSgrad   minHev minREev
Clayton unadj   1.5e+00 -6.1e+00     ---
Clayton adj     1.5e+00 -6.1e+00 9.8e-03
Plackett unadj  4.1e+02 -5.2e+00     ---
Plackett adj    4.1e+02 -5.2e+00 8.7e-03
Hougaard unadj  1.4e+01  7.7e-01     ---
Hougaard adj    1.4e+01  7.7e-01 7.7e-03
PoissonT        1.3e-05  1.3e+02 6.3e-12
PoissonI        2.0e-05  6.8e+01     ---
PoissonTI       7.1e-06  6.7e+01 2.0e-02
PoissonTIa      5.0e-05  9.4e+07 1.0e-01
```

5. Prediction of the treatment effect

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

```
R>   predict(allSurroRes, models = 'PoissonTI')
```

```
Treatment effect prediction for surrosurv object

   Poisson TI
                            1     2     3     4     5     6
    Treatment effects on S: -0.52 -0.42 -0.38 -0.08 -0.51 -0.38 ...
    Treatment effects on T: -0.26 -0.08 -0.27  0.41 -0.41 -0.15 ...
```

This function returns an object of class `predictSurrosurv`.

The predicted treatment effects can also be vizualied graphically using the linear regression of the effect on T given the effect on S. 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. reffig:predictions) can be obtained as follows:

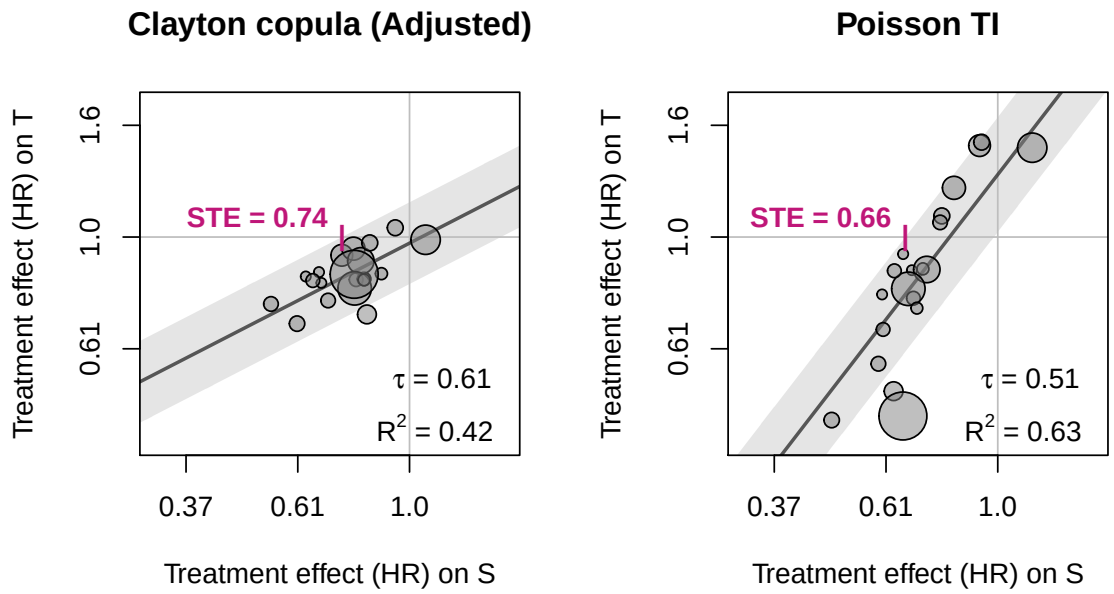

Figure 2: Predictions for the advanced GASTRIC meta-analysis (GASTRIC group 2013), as computed by the adjusted Clayton copula model, which had poor convergence metrics, and by the Poisson TI model, which was deemed to have converged. HR = hazard ratio.

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

The argument `surro.stats` controls whether the estimated Kendall's τ 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 (Burzykowski and Buyse 2006). 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`):

```
R>   ste(allSurroRes)
```

```
                beta   HR
Clayton.unadj  -0.61 0.54
Clayton.adj    -0.30 0.74
Plackett.unadj -0.61 0.54
Plackett.adj   -0.30 0.74
Hougaard.unadj -0.61 0.54
Hougaard.adj   -0.28 0.76
PoissonT       -0.12 0.88
PoissonTI      -0.41 0.66
PoissonTIa     -1.04 0.36
```

5.1. 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 and to cross-validate the results of the model fitted on $N-1$ trials thanks to the prediction in the left-out trial of the treatment effect on T, based on the observed effect on S (Michiels *et al.* 2009; Mauguen *et al.* 2013; Rotolo *et al.* 2017). The function `loovc()` 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, a fortiori the cross-validation is too. For that reason, the function `loovc()` 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 minimum between N and the maximum number of cores on the machine.

```
R>   loocvRes <- loocv(gastadv, models = c('Clayton', 'PoissonTI'))

Parallel computing on 8 cores (the total number of cores detected)
```

The results of the crossvalidation can be easily printed

```
R>   loocvRes


   Clayton copula (Unadjusted)
        1     2     3     4     5     6
obsBeta -0.31 -0.21 -0.09 -0.02 -0.22 -0.34 ...
lwr     -0.76 -0.65 -0.42 -0.51 -0.48 -0.62 ...
upr     -0.05  0.02  0.28  0.17  0.21  0.09 ...

   Clayton copula (Adjusted)
        1      2      3      4      5      6
obsBeta -0.309 -0.212 -0.095 -0.023 -0.222 -0.342 ...
lwr     -0.571 -0.491 -0.277 -0.358 -0.332 -0.448 ...
upr     -0.213 -0.130  0.105 -0.001  0.042 -0.078 ...

   Poisson TI
        1     2     3     4     5     6
obsBeta -0.31 -0.21 -0.09 -0.02 -0.22 -0.34 ...
lwr     -0.87 -0.67 -0.42 -0.59 -0.29 -0.54 ...
upr     -0.45 -0.21  0.57  0.23  0.23 -0.22 ...
```

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:

```
R>   plot(loocvRes)
```

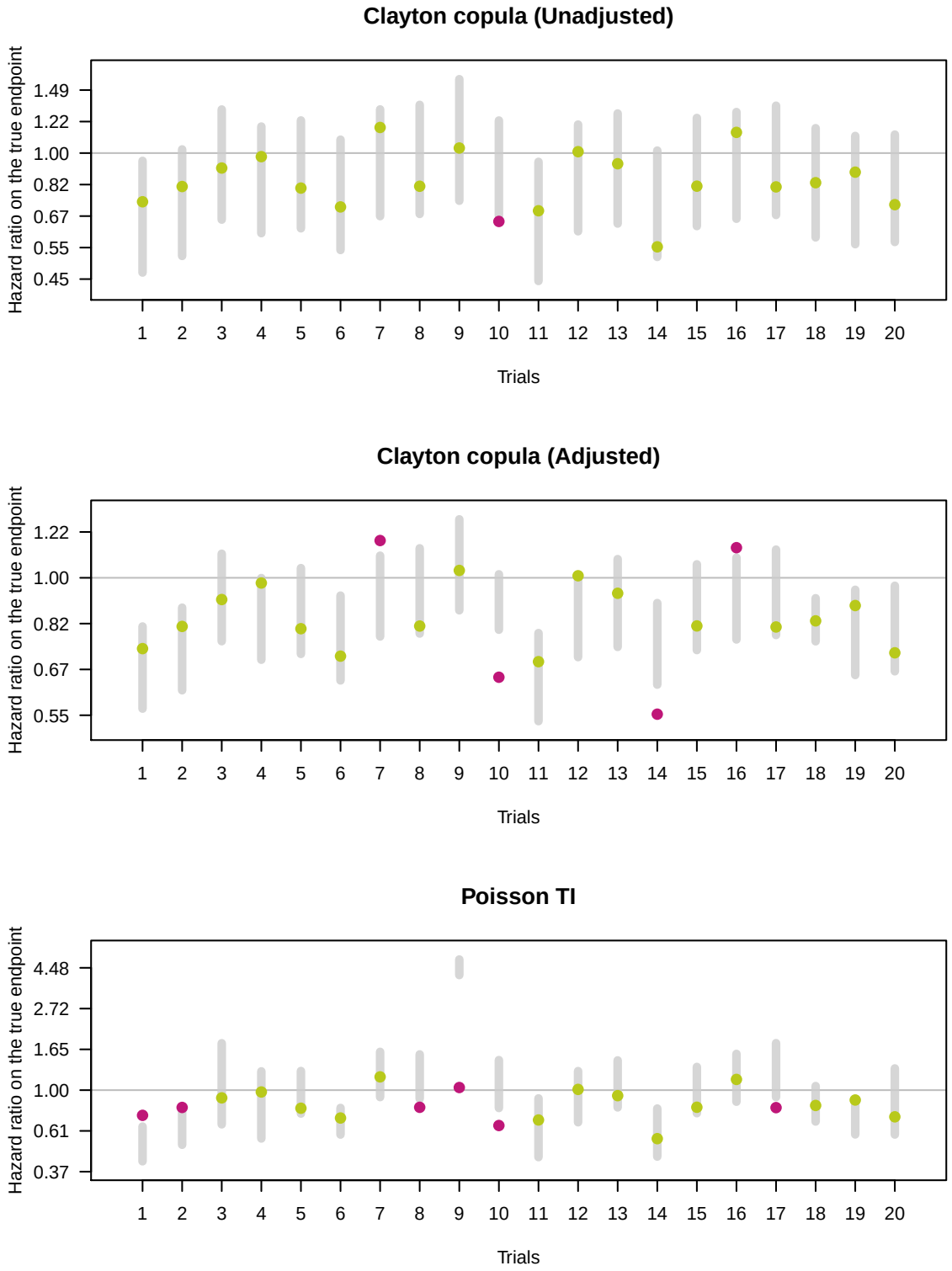

Figure 3: Leave-one-trial-out cross-validation results for the advanced GASTRIC meta-analysis (GASTRIC group 2013). Vertical lines are the 95% prediction intervals (PI) of the treatment effect on overall survival (OS). Dots are the observed treatment effects on OS (green = within the PI, magenta = out of the PI).

6. Utilities for data simulation

Most of the publications that discuss statistical methods for evaluating failure time surrogate endpoints do not show simulations till date. To our knowledge, rare exceptions are those papers by Burzykowski and Cortiñas Abrahantes (2005); Shi *et al.* (2011); Renfro *et al.* (2012, 2014, 2015).

6.1. Data generation based on a Clayton copula

The data geration method used by Burzykowski and Cortiñas Abrahantes (2005) and by Renfro *et al.* (2014, 2015) 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.

$$S_{ij} = -\log(U_{Sij})/\lambda_{Sij}, \quad \text{with } \lambda_{Sij} = \exp(\mu_S + m_{S_i} + \alpha_i Z_{ij}) \text{ and } U_{Sij} \sim U(0,1)$$

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

$$\begin{aligned} T_{ij} = -\log(U'_{Tij})/\lambda_{Tij}, \quad \text{with } & \lambda_{Tij} = \exp(\mu_T + m_{T_i} + \beta_i Z_{ij}), \\ & U'_{Tij} = \left[\left(U_{Tij}^{-\theta/(1+\theta)} - 1 \right) U_{Sij}^{-\theta} + 1 \right]^{-1/\theta}, \text{ and} \\ & U_{Tij} \sim U(0,1). \end{aligned}$$

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

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

The data geration method used by Shi *et al.* (2011) and by Renfro *et al.* (2012) is based on the results by Cowles (2004), 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, implemented in the function `simData.mx()`, which 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

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

- unit rate parameter exponential random variables Λ_{Sij} and Λ_{Tij} are generated from $-\log(U_{Sij})_{Sij}$ and $-\log(U_{Tij})$, with $U_{Sij} \sim U(0,1)$ and $U_{Tij} \sim U(0,1)$

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

$$\begin{aligned} S_{ij} &= \left(Y_{ij}^* \sqrt{2\Lambda_{Sij}} \right) \exp(\mu_S + m_{S_i} + \alpha_i Z_{ij}), \\ T_{ij} &= \left(Y_{ij}^* \sqrt{2\Lambda_{Tij}} \right) \exp(\mu_S + m_{T_i} + \alpha_i Z_{ij}). \end{aligned}$$

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

6.3. Data generation based on mixed proportional hazard models

In Rotolo *et al.* (201x), we also generated data using individual random effects to control individual-level surrogacy. 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 σ^2 depending on the scenario (according to the Kendall's τ)

- 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 Robert and Casella 2009, § 2.1.2)

$$\begin{aligned} S_{ij} &= -\log(U_{Sij})/\lambda_{Sij}, \quad \text{with } \lambda_{Sij} = \exp(\mu_S + m_{S_i} + \alpha_i Z_{ij} + u_{ij}) \text{ and } U_{Sij} \sim U(0,1), \\ T_{ij} &= -\log(U_{Tij})/\lambda_{Tij}, \quad \text{with } \lambda_{Tij} = \exp(\mu_T + m_{T_i} + \beta_i Z_{ij} + u_{ij}) \text{ and } U_{Tij} \sim U(0,1). \end{aligned}$$

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

Acknowledgments

The present work has been supported by the French National Cancer Institute (Institut National du Cancer, INCa), Grant SHS 2014-141, and by the Ligue Nationale Contre le Cancer.

References

Burzykowski T, Buyse M (2006). "Surrogate Threshold Effect: An Alternative Measure for Meta-Analytic Surrogate Endpoint Validation." *Pharmaceutical Statistics*, **5**(3), 173–186. doi:10.1002/pst.207.

Burzykowski T, Cortiñas Abrahantes J (2005). "Validation in the Case of Two Failure-Time Endpoints." In T Burzykowski, G Molenberghs, M Buyse (eds.), *The Evaluation of Surrogate Endpoints*, pp. 163–194. Springer, New York, NY.

Burzykowski T, Molenberghs G, Buyse M (2006). *The Evaluation of Surrogate Endpoints*. Springer Science & Business Media, New York, NY.

Burzykowski T, Molenberghs G, Buyse M, Geys H, Renard D (2001). "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), 405–422. doi:10.1111/1467-9876.00244.

Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H (2000). "The Validation of Surrogate Endpoints in Meta-Analyses of Randomized Experiments." *Biostatistics*, **1**(1), 49–67. doi:10.1093/biostatistics/1.1.49.

Buyse M, Molenberghs G, Paoletti X, Oba K, Alonso A, Van der Elst W, Burzykowski T (2016). "Statistical Evaluation of Surrogate Endpoints with Examples from Cancer Clinical Trials." *Biometrical Journal*, **58**(1), 104–132. doi:10.1002/bimj.201400049.

Clayton D (1978). "A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence." *Biometrika*, **65**, 141–151. doi:10.1093/biomet/65.1.141.

Cowles M (2004). "Evaluating Surrogate Endpoints for Clinical Trials: A Bayesian Approach." *Technical report*, University of Iowa. URL https://stat.uiowa.edu/sites/stat.uiowa.edu/files/techrep/tr334.pdf.

Crowther MJ, Riley RD, Staessen JA, Wang J, Gueyffier F, Lambert PC (2012). "Individual Patient Data Meta-Analysis of Survival Data Using Poisson Regression Models." *BMC Medical Research Methodology*, **12**(1), 34. doi:10.1186/1471-2288-12-34.

Duchateau L, Janssen P (2008). *The Frailty Model*. Springer Verlag, New York, NY. ISBN 978-0-387-72834-6. doi:10.1007/978-0-387-72835-3.

Gasparrini A (2015). "**mvmeta**: Multivariate and Univariate Meta-Analysis and Meta-Regression." R package version 0.4-7. URL https://CRAN.R-project.org/package=mvmeta.

Gasparrini A, Armstrong B, Kenward M (2012). "Multivariate Meta-Analysis for Non-Linear and Other Multi-Parameter Associations." *Statistics in Medicine*, **31**, 3821–3839. doi:10.1002/sim.5471.

GASTRIC group (2013). "Role of Chemotherapy for Advanced/Recurrent Gastric Cancer: An Individual-Patient-Data Meta-Analysis." *European Journal of Cancer*, **49**(7), 1565–1577. doi:10.1016/j.ejca.2012.12.016.

Goutis C, Casella G (1999). "Explaining the Saddlepoint Approximation." *The American Statistician*, **53**(3), 216–224. doi:10.1080/00031305.1999.10474463.

Hofert M, Kojadinovic I, Maechler M, Yan J (2016). "**copula**: Multivariate Dependence with Copulas." R package version 0.999-15. URL https://CRAN.R-project.org/package=copula.

Hougaard P (1986). "A Class of Multivariate Failure Time Distributions." *Biometrika*, **73**, 671–678. doi:10.1093/biomet/73.3.671.

Kendall MG (1938). "A New Measure of Rank Correlation." *Biometrika*, **30**(1/2), 81–93. URL http://www.jstor.org/stable/2332226.

Kovalchik S (2013). "Re: Exponential Proportional Hazard Model." http://r.789695.n4.nabble.com/exponential-proportional-hazard-model-td805536.html. Accessed: 2016-11-14.

Laird N, Olivier D (1981). "Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques." *Journal of the American Statistical Association*, **76**(374), 231–40. doi:10.1080/01621459.1981.10477634.

Mauguen A, Pignon JP, Burdett S, Domerg C, Fisher D, Paulus R, Mandrekar SJ, Belani CP, Shepherd FA, Eisen T, Pang H, Collette L, Sause WT, Dahlberg SE, Crawford J, O'Brien M, Schild SE, Parmar M, Tierney JF, Le Pechoux C, Michiels S (2013). "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), 619–26. doi:10.1016/S1470-2045(13)70158-X.

Michiels S, Le Maître A, Buyse M, Burzykowski T, Maillard E, Bogaerts J, Vermorken JB, Budach W, Pajak TF, Ang KK, Bourhis J, Pignon JP (2009). "Surrogate endpoints for overall survival in locally advanced head and neck cancer: meta-analyses of individual patient data." *The Lancet Oncology*, **10**(4), 341–50. doi:10.1016/S1470-2045(09)70023-3.

Munda M, Rotolo F, Legrand C (2012). "**parfm**: Parametric Frailty Models in R." *Journal of Statistical Software*, **51**(1), 1–20. doi:10.18637/jss.v051.i11.

Nash JC (2012). "On Best Practice Optimization Methods in R." *Journal of Statistical Software*, **60**(2), 1–14. doi:10.18637/jss.v043.i09.

Nash JC, Varadhan R, Grothendieck G (2013). "**opitmx**: A Replacement and Extension of the `optim()` Function." R package version 2013.8.7. URL https://CRAN.R-project.org/package=optimx.

Oba K, Paoletti X, Alberts S, Bang YJ, Benedetti J, Bleiberg H, Catalano P, Lordick F, Michiels S, Morita S, Ohashi Y, Pignon JP, Rougier P, Sasako M, Sakamoto J, Sargent D, Shitara K, Van Cutsem E, Buyse M, Burzykowski T (2013). "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), 1600–7. doi:10.1093/jnci/djt270.

Paoletti X, Oba K, Bang YJ, Bleiberg H, Boku N, Bouché O, Catalano P, Fuse N, Michiels S, Moehler M, Morita S, Ohashi Y, Ohtsu A, Roth A, Rougier P, Sakamoto J, Sargent D, Sasako M, Shitara K, Thuss-Patience P, Van Cutsem E, Burzykowski T, Buyse M (2013). "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), 1667–1670. doi:10.1093/jnci/djt269.

Plackett RL (1965). "A class of Bivariate Distributions." *Journal of the American Statistical Association*, **60**, 516–522. doi:10.1080/01621459.1965.10480807.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Renfro LA, Shang H, Sargent DJ (2015). "Impact of copula directional specification on multi-trial evaluation of surrogate endpoints." *Journal of Biopharmaceutical Statistics*, **25**, 857–877. doi:10.1080/10543406.2014.920870.

Renfro LA, Shi Q, Sargent DJ, Carlin BP (2012). "Bayesian adjusted R^2 for the meta-analytic evaluation of surrogate time-to-event endpoints in clinical trials." *Statistics in Medicine*, **31**(8), 743–761. doi:10.1002/sim.4416.

Renfro LA, Shi Q, Xue Y, Li J, Shang H, Sargent DJ (2014). "Center-within-trial versus trial-level evaluation of surrogate endpoints." *Computational Statistics & Data Analysis*, **78**, 1–20. doi:10.1016/j.csda.2014.03.011.

Robert C, Casella G (2009). *Introducing Monte Carlo Methods with R*. Springer Science & Business Media. doi:10.1007/978-1-4419-1576-4.

Rotolo F (2016). "**surrosurv**: Evaluation of Failure Time Surrogate Endpoints in Individual Patient Data Meta-Analyses." R package version 1.1.4. URL https://CRAN.R-project.org/package=surrosurv.

Rotolo F, Munda M (2015). "**parfm**: Parametric Frailty Models." R package version 2.5.10. URL https://CRAN.R-project.org/package=parfm.

Rotolo F, Paoletti X, Buyse M, Burzykowski T, Michiels S (201x). "Evaluation of Failure Time Surrogate Endpoints in Individual Patient Data Meta-Analyses of Randomized Clinical Trials: a Poisson Approach." *Under Review*.

Rotolo F, Pignon JP, Bourhis J, Marguet S, Leclercq J, Ng WT, Ma J, Chan A, Huang6 PY, Zhu G, Chua D, Chen Y, Mai HQ, Kwong DLW, Soong YL, Moon J, Tung Y, Chi KH,

Fountzilas G, Zhang L, Pun Hui E, Lee AWM, Blanchard P, Michiels S (2017). "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.

Shi Q, Renfro LA, Bot BM, Burzykowski T, Buyse M, Sargent DJ (2011). "Comparative assessment of trial-level surrogacy measures for candidate time-to-event surrogate endpoints in clinical trials." *Computational Statistics & Data Analysis*, **55**(9), 2748–2757. doi:10.1016/j.csda.2011.03.014.

van Houwelingen HC, Arends L, Stijnen T (2002). "Advanced Methods in Meta-Analysis: Multivariate Approach and Meta-Regression." *Statistics in Medicine*, **21**, 589–624. doi:10.1002/sim.1040.

Whitehead J (1980). "Fitting Cox's Regression Model to Survival Data Using GLIM." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, pp. 268–275. doi:10.2307/2346901.

Yan J (2007). "Enjoy the Joy of Copulas: With a Package **copula**." *Journal of Statistical Software*, **21**(1), 1–21. doi:10.18637/jss.v021.i04.

A. Data poissonization

Fitting auxiliary Poisson models for estimating the parameters of a proportional hazard model (Whitehead 1980; Crowther *et al.* 2012) needs that data are rearranged in order to provide, for each time period, the number of events and the total time passed at risk. The function `poissonize()`in the **surrosurv** package allows to easily perform the necesasry data manipulaton easily and quickly. The core of the function has been derived from the original code publicly shared by Kovalchik (2013).

The main argument of the `poissonize()` function is `data`, a data frame with columns: `id`, the patient identifyier; `time`, the event/censoring time; `status`, the event (`1`) or censoring (`0`) indicator; `...`, other factors such like the covariables needed in the regression model.

The breakpoints between time intervals can be entered in the second argument, `all.breaks`. Otherwise, if `all.breaks` is not specified, one can specify either the width of the time intervals `interval.width`, or their number `nInts` (used only also if `is.null(interval.width)`).

Any other variables to be kept in the poissonized data frame can be entered in `factors`. The last argument (`compress`) is a logical indicating whether the record with the same factor profile should be summarized into one record, i. e. whether the data should be expressed in a short form.

In the advanced GASTRIC cancer example, we first change the column names in order to match the ones needed by `poissonize()`:

```
R>   gastadv.poi <- gastadv
R>   gastadv.poi$time <- gastadv.poi$timeT / 365.25
R>   gastadv.poi$status <- gastadv.poi$statusT
```

We fit the proportional hazard model, to which we will compare the results of the auxiliary Poisson model.

```
R>   fitcox <- coxph(Surv(time, status) ~ trt, data = gastadv.poi)
R>   cox.base <- basehaz(fitcox, centered = FALSE)
```

We 'possonize' the data over 10 intervals (the default) and we fit the auxiliary Poisson model.

```
R>   gastadv.poi <- poissonize(gastadv.poi, nInts = 10, factors = 'trt')
R>   gastadv.poi

            interval  trt   m  Rt    N
1                  0 -0.5 181 292 1668
2    0.1832128678987 -0.5 180 173 1475
3   0.30921697467488 -0.5 192 149 1288
4  0.435221081451061 -0.5 159 132 1088
5  0.567018480492813 -0.5 154 114  912
6  0.703885010266941 -0.5 156 108  751
7  0.867545516769336 -0.5 157 103  584
8   1.07320739219713 -0.5 143 101  414
9   1.39328678986995 -0.5 117  97  239
10  2.07255030800821 -0.5  60  87   94
11                 0  0.5 216 421 2401
12   0.1832128678987  0.5 221 258 2167
13  0.30921697467488  0.5 213 229 1935
14 0.435221081451061  0.5 247 207 1706
15 0.567018480492813  0.5 237 181 1446
16 0.703885010266941  0.5 225 176 1203
17 0.867545516769336  0.5 228 171  965
18  1.07320739219713  0.5 221 183  715
19  1.39328678986995  0.5 211 205  460
20  2.07255030800821  0.5 117 171  204

R>   fitpoi <- glm(m ~ -1 + interval + trt + offset(log(Rt)),
+                 data = gastadv.poi, fam = 'poisson')
```

The function `plotsson()` can be used to draw the survival curves (or the instantaneous hazard) estimated by the auxiliary Poisson model:

```
R>   plot(stepfun(cox.base$time[-nrow(cox.base)],
+                exp(-cox.base$hazard)),
+        ylim = 0:1, xlim = c(0, 5), col = 1, lwd = 2, bty = 'l', yaxs = 'i',
+        do.points = FALSE, verticals = FALSE, xaxs = 'i',
+        main = 'Overall Survival\nAdvanced GASTRIC meta-analysis',
+        xlab = 'Years', ylab = 'Survival probability')
R>   lines(stepfun(cox.base$time[-nrow(cox.base)],
+                exp(-cox.base$hazard * exp(coef(fitcox)['trt']))),
+        col = 2, pch = '', lwd = 2)
```

The treatment effect estimated by the Cox model is -0.14 (SE $= 0.03$), and it is of -0.14 (SE $= 0.03$) when using the auxiliary Poisson model.

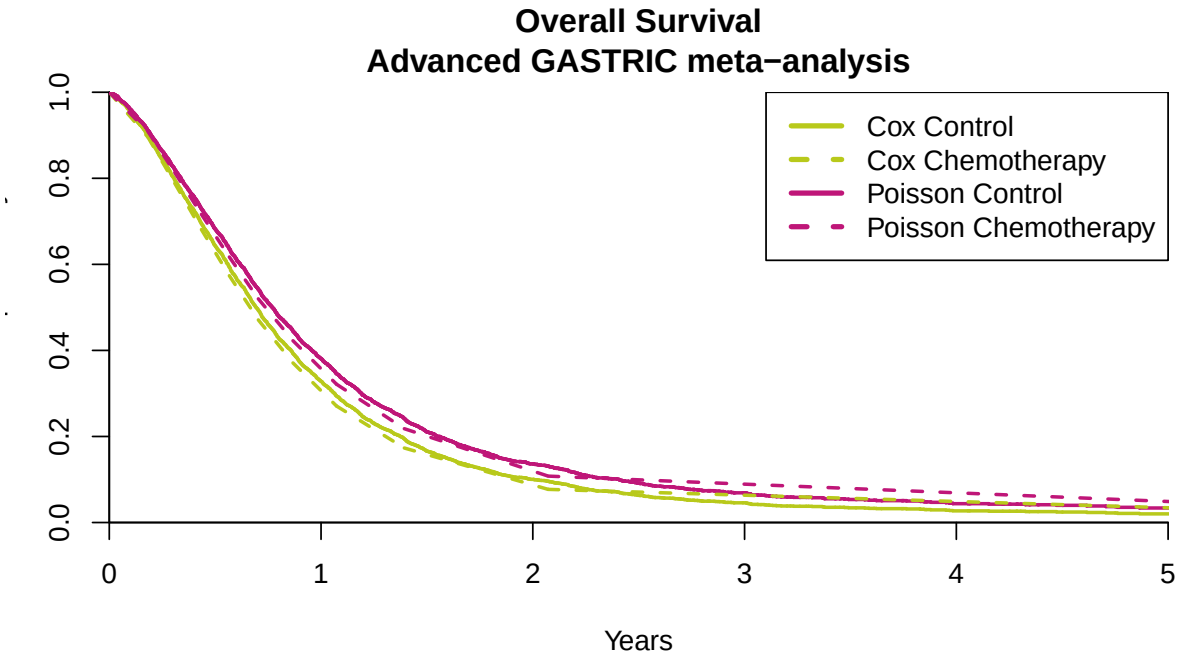

Figure 4: Overall survival curves in the advanced GASTRIC meta-analysis (GASTRIC group 2013). **(a)** Comparison between the survival probability obtanied using the Breslow estimator in the Cox model (solid lines) and those obtained using the auxiliary Poisson model (dashed lines). **(b)** Piecewise constant hazard estimated by the auxiliary Poisson model

Affiliation:

Federico Rotolo, Xavier Paoletti, and Stefan Michiels

Service de Biostatistique et d'Épidémiologie, Gustave Roussy
INSERM U1018 OncoStat, CESP
114, Rue Edouard Vaillant – Villejuif, F-94805, France
E-mail: federico.rotolo@gustaveroussy.fr