QoLR: An R Package for the Longitudinal Analysis of Health-Related Quality of Life in Oncology

Amélie Anota
University Hospital of Besançon

Marion Savina
Bergonie Institute of Bordeaux

Caroline Bascoul-Mollevi
Cancer Institute of Montpellier

Franck Bonnetain
University Hospital of Besançon

Abstract

Health-related quality of life has become increasingly important in clinical trials over the past two decades. The R package **QoLR** is a recently developed package for the longitudinal analysis of health-related quality of life in oncology. This package contains the scoring of most of the European Organisation for Research and Treatment of Cancer quality of life questionnaires and some programs to analyze the time to a health-related quality of life score deterioration as a modality of longitudinal analysis in oncology.

Keywords: health-related quality of life, longitudinal analysis, time to deterioration, R package **QoLR**.

1. Introduction

1.1. Health-related quality of life

Health-related quality of life (HRQOL) is a subjective clinical endpoint that has become increasingly important in clinical trials in oncology over the past two decades (Osoba 2011). Although overall survival is still considered as the primary objective and the primary endpoint in many studies, most clinical trials now integrate HRQOL as an endpoint in order to investigate the clinical benefit for the patient. It even seems that HRQOL could become the primary endpoint in metastatic settings.

1.2. Assessment

HRQOL is mainly studied through validated questionnaires filled out by the patients at several timepoints during the study. In oncology, the European Organisation for Research and Treatment of Cancer (EORTC) has developed a core questionnaire, namely the QLQ-C30, to assess HRQOL in cancer patients with 30 questions (items) and some supplementary modules for disease-specific treatment measurements (Aaronson *et al.* 1993).

The EORTC HRQOL questionnaires are composed of a set of items with several categories of responses, usually constructed on a Likert scale (e.g., "Not at all"/"a little"/"quite a bit"/"very much", coded 1/2/3/4). These questions allow the estimation of one or more HRQOL dimensions through the calculation of scores. Generally, a score is the weighted sum of patients' items responses regarding one dimension and can be calculated if at least half of the items are answered. The scores are often normalized on a 0–100 scale in order to facilitate comparison with one or more supplementary HRQOL questionnaires.

For most of the HRQOL questionnaires, a high score corresponds to a high level of functioning for a functional scale, and to high presence of symptoms for a symptomatic scale. Conversely, a low score corresponds to a low level of functioning for a functional scale and to low presence of symptoms for a symptomatic scale. In this way, patients with a high level of HRQOL should have high scores on functional scales and low scores on symptomatic scales. For questionnaires that include an item about global health or global HRQOL (as for the EORTC QLQ-C30), the score obtained is coded like a functional scale.

The method to calculate the scores of the EORTC HRQOL questionnaire is defined in the EORTC scoring manual (Fayers, Aaronson, Bjordal, Curran, and Grønvold 1999). Briefly, all the scores of the EORTC HRQOL questionnaires are obtained with the same procedure:

Let $I_1, \ldots, I_n$ be the n items answers to the studied dimension.

The first step, common to both functional and symptomatic scales, is the calculation of a raw score (RS), i.e., the mean of the items:

$$\text{RS} = \frac{I_1 + \cdots + I_n}{n} \tag{1}$$

If there are some missing items, the denominator equals the number of non-missing items. If more than half the items are missing, the raw score cannot be calculated and then the dimension score is missing.

The second step is a linear transformation of the scores to a 0–100 scale to obtain a score S:

- for functional scales:

$$S = \left(1 - \frac{\text{RS} - 1}{\text{range}}\right) \times 100 \tag{2}$$

- for symptomatic scales:

$$S = \left(\frac{\text{RS} - 1}{\text{range}}\right) \times 100 \tag{3}$$

- for global health status:

$$S = \left(1 - \frac{\text{RS} - 1}{\text{range}}\right) \times 100 \tag{4}$$

1.3. Longitudinal analysis

The longitudinal analysis of HRQOL is a major challenge due to different parameters that have to be taken into account in the analysis.

Firstly, the longitudinal assessment of HRQOL may be compromised by the presence of missing data. Two types of missing data can occur: intermittent missing data (e.g., if a patient forgot to complete one or more questions at one measurement time) or monotone missing data (e.g., if a patient dropped out before the end of the study). The presence of missing data can be informative or non-informative of patient's health status. For example, if a patient dropped out the study due to disease progression, monotone missing data occur and can provide information regarding the patient's HRQOL level: the patient's HRQOL level is likely to have decreased. Thus, missing data are missing not at random. Regarding intermittent missing data, in some circumstances, we can suppose that the patient's HRQOL level remains unchanged between two available measures (missing at random or completely at random).

Secondly, self-assessment of HRQOL is subjective, i.e., it is dependent on the patient's internal standards and definition of HRQOL (Wiklund 2004; Bullinger 2002; Ubel, Peeters, and Smith 2010). However, patients' adaptation to treatment toxicity and their acceptance of the disease may mean that they do not necessarily assess their HRQOL with the same criteria at all measurement times. These changes can be reflected by a response shift effect (Gibbons 1999; Sprangers and Schwartz 1999; Korfage, de Koning, and Essink-Bot 2007) and can bias the interpretation of longitudinal HRQOL analysis if not taken into account in the study design or at least in the analysis method (Ahmed, Mayo, Corbiere, Wood-Dauphinee, Hanley, and Cohen 2005).

Response shift is defined as "a change in the meaning of one's self-evaluation of a target construct as a result of: (a) a change in the respondent's internal standards of measurement (i.e., scale recalibration); (b) a change in the respondent's values, i.e., the importance of component domains constituting the target construct) or (c) a redefinition of the target construct (i.e., reconceptualization)" (Sprangers and Schwartz 1999). In this way, the choice of the level of HRQOL reference to qualify a change like deterioration can be a major concern: the baseline score is not systematically the reference score.

Finally, the longitudinal analysis of HRQOL involves the question of the minimal clinically important difference (MCID). Indeed, there is a difference between the notion of "statistically significant" and "clinically significant". The MCID was defined by Jaeschke, Singer, and Guyatt (1989) as the smallest change/difference in HRQOL scores perceived as clinically important. Apart from statistical significance alone, the MCID is a key component that makes it possible to judge the clinical meaning of the results. Terwee *et al.* (2010) have reported that many methods have been proposed to determine the MCID, but no standard has been proposed at this time. These methods are generally grouped in two categories: the anchor-based and distribution-based methods. The anchor-based methods focus on the relation between the HRQOL scores and an external criterion that has clinical relevance (Lydick and Epstein 1993). This anchor can be a criterion of disease progression, for example but can also be a patient's self-assessment of the global HRQOL change (Jaeschke *et al.* 1989). Distribution-based methods are based on the distribution of the HRQOL scores on the obtained sample. For example, the MCID can then be defined as the half of the standard error (Norman, Sloan, and Wyrwich 2003) or the standardized response mean (Wyrwich, Nienaber, Tierney, and

Wolinsky 1999a; Wyrwich, Tierney, and Wolinsky 1999b).

For the EORTC HRQOL questionnaires, Osoba, Rodrigues, Myles, Zee, and Pater (1998) have demonstrated that a MCID of 5 points is a small change, a change between 10 and 20 points is moderate and more than 20 points indicates very much change. These thresholds are used for illustration purposes but they are not a gold standard. Moreover, the MCID can also be questionnaire-specific. More recent and scale-specific guidelines are available for clinical significance for the EORTC QLQ-C30 (Cocks, King, Velikova, St-James, Fayers, and Brown 2011).

Due to the complexity of the longitudinal assessment of HRQOL, there is still no gold standard to analyze HRQOL over time in oncology. Moreover, another challenge of statistical methods for analyzing longitudinal HRQOL is to propose meaningful results for the clinicians. There is a need to develop statistical methods adapted for the decision-makers. Longitudinal results should have the ability to translate findings into information that decision makers find understandable and compelling. Thus, different methods to analyze longitudinal HRQOL have been proposed (Pan, Chen, Chung, Wang, Chen, and Hsiung 2012; Hunger, Döring, and Holle 2012; Penar-Zadarko, Binkowska-Bury, Wolan, Gawelko, and Urbanski 2012; Cnaan, Laird, and Slasor 1997). The most widely used is the linear mixed model (Diggle 1988). Survival analysis approaches, like the time to deterioration in a HRQOL score, have recently been proposed as a modality of longitudinal HRQOL analysis in cancer patients (Bonnetain *et al.* 2010; Hamidou *et al.* 2011).

The linear mixed model is optimal for a study design with 2 to 5 measurement times (Fairclough 2010). In this model, time is considered as a categorical variable. Moreover, these models are only adapted for studies whose HRQOL assessments are performed in some periods with few amplitude within patients. In case of a missing not at random profile for missing data, a pattern mixture model should be applied in order to produce robust results (Pauler, McCoy, and Moinpour 2003). However, these sub models are rarely applied mainly because of the complexity of the construction of the patterns. Moreover, at this time, these models do not deal with the occurrence of a response shift effect. Finally, these models can not provide results easy to understand for the clinician who is not familiar with the beta change and the mixed model approach.

Contrary to the linear mixed model, "time to deterioration" (TTD) models can propose clinically meaningful results with hazard ratios and log-rank tests (Bonnetain *et al.* 2010; Hamidou *et al.* 2011). Moreover, these models can handle the presence of missing data: when intermittent missing data occur for one patient, we can consider that the patient's HRQOL level remains unchanged since the previous available measure; whereas monotone missing data can be due to a deterioration of the patient's health status in case of advanced/metastatic setting. Finally, these models can take into account the occurrence of the recalibration component of the response shift effect by choosing different scores as the reference score to qualify the deterioration (Anota *et al.* 2013).

The aim of this paper is to present the R (R Core Team 2017) package **QoLR** which is available from the Comprehensive R Archive Network at `https://CRAN.R-project.org/package=QoLR`. This package makes it possible to calculate the scores of the EORTC QLQ-C30 and most of its modules and to determine the time to deterioration in a HRQOL score as a modality of longitudinal analysis.

2. Time to deterioration definitions

To date, several definitions of TTD in a HRQOL score have been proposed depending on event definition and censoring rules. The event definition can be defined according to the reference score, the MCID, missing scores, including death or not. The most intuitive definition is the time from inclusion in the study to a first deterioration of the score with a MCID of at least k points as compared to the baseline score (Hamidou *et al.* 2011). Patients with no deterioration before drop-out from the study are censored at the time of last HRQOL questionnaire completion or last follow-up. According to the construction of the score, deterioration corresponds to an increase (e.g., for symptomatic scales of the EORTC questionnaires) or a decrease (e.g., for functional scales of the EORTC questionnaires) of the score. Between two available HRQOL scores, the level of HRQOL is supposed to be constant.

The notion of "deterioration" requires a reference score. Generally, the reference score is the baseline score (before randomization or at inclusion in the absence of randomization). However, in order to take into account the occurrence of a response shift effect, the reference can also be

- the best level of HRQOL already experienced by the patient (i.e., the best previous score);

- or the previous score HRQOL score for the patient (i.e., immediately preceding score).

The value of these scores can change over time according to the patient's experience of treatment and disease progression. Using these changing references instead of baseline measurement could be considered as an alternative way to take into account the occurrence of the recalibration component of the response shift effect.

Furthermore, for event definition, we can consider deterioration as definitive (i.e., absorbing state) or not depending on the therapeutic setting. This induces two concepts: the TTD and the time until definitive deterioration (TUDD). Several definitions of TUDD have been proposed, according to the notion of "definitive deterioration". The TUDD has been defined as:

1. the time from inclusion in the study to a first deterioration of at least k points as compared to the reference score,

    (a) with no further improvement of more than k points as compared to the reference score,

    (b) or if the time of the deterioration correspond to the last observed HRQOL score, thus the patient dropped out (i.e., no more HRQOL data available) just after the deterioration was observed resulting in missing data (Bonnetain *et al.* 2010).

2. the time from inclusion in the study to a first deterioration of at least k points as compared to the reference score,

    (a) thereafter maintaining this deterioration of at least k points for all following scores, i.e., the deterioration is observed for all the scores following the first deterioration,

    (b) or if the patient dropped out just after the deterioration was observed resulting in missing data.

3. the time from inclusion in the study to a first deterioration of at least k points as compared to the reference score,

    (a) with no further improvement of more than k points as compared to the score qualifying the deterioration (i.e., the score at the time of the first deterioration observed),

    (b) or if the patient dropped out just after the deterioration was observed resulting in missing data.

To illustrate this, the following equations correspond to the three definitions of TUDD of a X score with a k-point MCID observed at time i as compared to the reference score X_{ref} assuming that X represents a functional scale (i.e., a deterioration is observed when the score decreases):

$\exists i > 1; X_i \leq X_{\mathrm{ref}} - k$ and:

1. $\forall j > i, X_j \leq X_{\mathrm{ref}} + k$ for the first definition of TUDD;

2. $\forall j > i, X_j \leq X_{\mathrm{ref}} - k$ for the second definition of TUDD;

3. $\forall j > i, X_j \leq X_i + k$ for the last definition of TUDD.

For example, if a patient has a reference score equal to 60 points for a functional score ($X_{\mathrm{ref}} = 60$), then a deterioration with a 5-point MCID as compared to the reference score is observed at time i if $X_i \leq 55$ ($X_i \leq X_{\mathrm{ref}} - 5$-point MCID). For example, let X_i equal 40. For each of the three definitions of TUDD previously outlined, this deterioration in time i is definitive if:

1. $\forall j > i, X_j \leq 65$ ($X_{\mathrm{ref}} + 5$-point MCID $= 60 + 5$) for the first definition of TUDD;

2. $\forall j > i, X_j \leq 55$ ($X_{\mathrm{ref}} - 5$-point MCID $= 60 - 5$) for the second definition of TUDD;

3. $\forall j > i, X_j \leq 45$ ($X_i + 5$-point MCID $= 40 + 5$) for the last definition of TUDD.

Then :

1. if a time k ($k > i$), $X_k = 66$ or more, then the deterioration observed at time i is not definitive according to the first definition of TUDD;

2. if a time k ($k > i$), $X_k = 56$ or more, then the deterioration observed at time i is not definitive according to the second definition of TUDD;

3. if a time k ($k > i$), $X_k = 46$ or more, then the deterioration observed at time i is not definitive according to the last definition of TUDD.

Table 1 indicates the event definition according to the definition of TTD or TUDD retained.

The time until definitive HRQOL score deterioration is mainly applied in advanced or palliative setting. All-cause death can also be considered as an event if the patient did not experience deterioration before death. In this way, TUDD or death could be redefined as "HRQOL deterioration-free survival".

Definition	Reference score X_ref at T_i				Event definition at T_i	
	Baseline	Best previous score		Immediately		
MCID $=$ k points	score	Event	Event	preceding	Event $=$ decrease	Event $=$ increase
		decrease	increase	score		
TTD $\geq k$ points	X_1	$\max_{j\leq i}(X_j)$	$\min_{j\leq i}(X_j)$	X_{i-1}	$T_i = \min_{j>1}(T_j),$ $X_j \leq X_\mathrm{ref} - k$	$T_i = \min_{j>1}(T_j),$ $X_j \geq X_\mathrm{ref} + k$
TUDD $\geq k$ points with no further improvement $> k$ points as compared to X_ref	X_1	$\max_{j\leq i}(X_j)$	$\min_{j\leq i}(X_j)$	X_{i-1}	$T_i = \min_{j>1}(T_j)/$ $\bullet X_j \leq X_\mathrm{ref} - k$ $\bullet \forall m > i, X_m \leq X_\mathrm{ref} + k$	$T_i = \min_{j>1}(T_j)/$ $\bullet X_j \geq X_\mathrm{ref} + k$ $\bullet \forall m > i, X_m \geq X_\mathrm{ref} - k$
TUDD $\geq k$ points with a deterioration observed at all times following time T_i	X_1	$\max_{j\leq i}(X_j)$	$\min_{j\leq i}(X_j)$	X_{i-1}	$T_i = \min_{j>1}(T_j)/$ $\bullet X_j \leq X_\mathrm{ref} - k$ $\bullet \forall m > i, X_m \leq X_\mathrm{ref} - k$	$T_i = \min_{j>1}(T_j)/$ $\bullet X_j \geq X_\mathrm{ref} + k$ $\bullet \forall m > i, X_m \geq X_\mathrm{ref} + k$
TUDD $\geq k$ points with no further improvement $> k$ points as compared to X_i	X_1	$\max_{j\leq i}(X_j)$	$\min_{j\leq i}(X_j)$	X_{i-1}	$T_i = \min_{j>1}(T_j)/$ $\bullet X_j \leq X_\mathrm{ref} - k$ $\bullet \forall m > i, X_m \leq X_i + k$	$T_i = \min_{j>1}(T_j)/$ $\bullet X_j \geq X_\mathrm{ref} + k$ $\bullet \forall m > i, X_m \geq X_i - k$

Table 1: Event definitions at time T_i according to the definition of time to score X deterioration as compared to the reference score X_ref.

Patients with no score available are excluded from the time-to-deterioration analyses. Patients with no baseline score are usually censored at baseline and those with no follow up scores but with a baseline score are censored one day after baseline. As for other analyses of HRQOL, in the TTD analyses, some sensitivity analyses could be performed:

- Considering patients with no baseline score as events;

- Considering patients with no follow-up score as events;

- Varying the MCID.

For example, let us choose a MCID of 10 points instead of the 5 points initially fixed. Regarding the TUDD, as $5 < 10$, then $\mathsf{P}(\mathrm{TUDD} \geq 5) \geq \mathsf{P}(\mathrm{TUDD} \geq 10)$, i.e., the probability that a patient presented a definitive deterioration with 5-point MCID is greater than the probability that a patient presented a definitive deterioration with a 10-point MCID.

Regarding the TUDD, if a patient experienced a definitive deterioration with a 10-point MCID but not with a 5-point MCID for any one of the proposed definitions, then this patient must be considered to have also presented a definitive deterioration with a 5-point MCID, and the event time will be the time of the 10-point deterioration. Indeed, for patients experiencing a
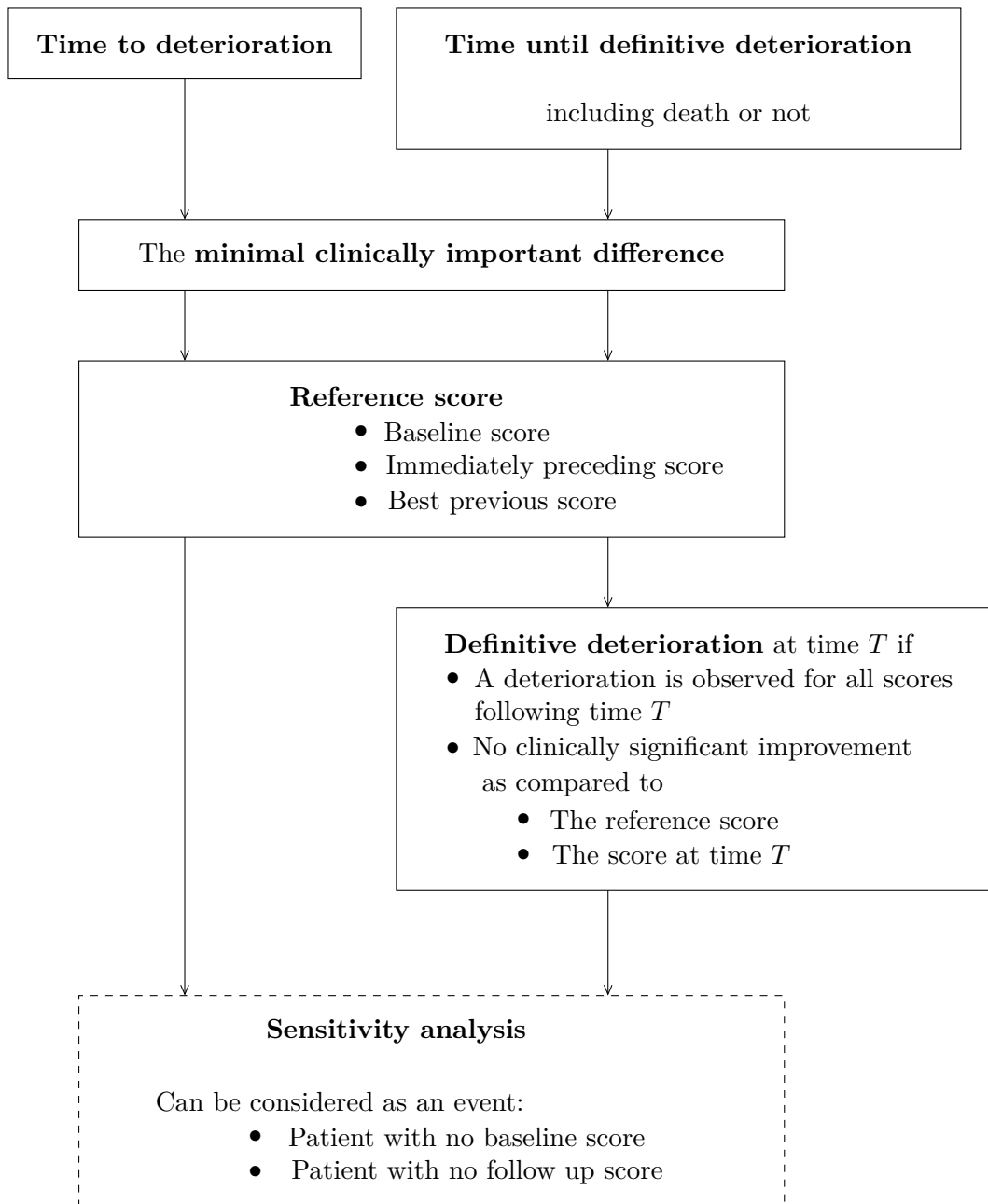
Figure 1: Flowchart of the choice of definition of time to a health-related quality of life score deterioration.

TUDD with both a 5-point MCID and a 10-point MCID, the TUDD with a 5-point MCID must be the time of the first deterioration observed (5-point or 10-point MCID). In this way, if the deterioration with a 10-point MCID occurs first, we have to impose that the time of the 5-point deterioration equals the time of the 10-point MCID deterioration.

Figure 1 summarizes the different proposed definitions of TTD and TUDD. To have a valid observation, we need a date of HRQOL measure and a HRQOL score. When the reference

score is the best previous score or the immediately preceding score, patients with no baseline score but with a post baseline score are kept in the TTD analyses if they have at least one follow up score available after the reference score. They are not censored at baseline. The first reference score is the first available score.

As TTD analyses belong to survival analyses, the TTD estimation can be calculated using the Kaplan-Meier or actuarial method and described using median and 95% confidence interval (CI). The Kaplan-Meier method is based on the intuitive idea that to be alive at time T, one has to be alive just before time T and not die at time T (Goel, Khanna, and Kishore 2010). Contrary to the Kaplan-Meier method, in the actuarial method probabilities are estimated for fixed time intervals, not determined by the date of observed death. Both methods can handle the presence of censored data, i.e., patients still alive at the end of the study.

In time to deterioration analyses, the event is "HRQOL score deterioration".

The Kaplan-Meier estimation is given by the following formula:

$$S(t) = \prod_{t_i \leq t} \frac{n_i - m_i}{n_i} \tag{5}$$

where $n_i = n_{i-1} - m_{i-1} - c_{i-1}$ and:

- n_i is the number of subjects at risk at time i, i.e., the number of patients still in the study and who do not present a deterioration until time $i - 1$;

- m_i is the number of events observed at time i, i.e., the number of patients experiencing a HRQOL score deterioration at time i;

- c_i is the number of censored patients at time i, i.e., the number of patients who drop out at time i and who did not experienced a HRQOL deterioration before.

TTD can then be compared according to treatment arm in case of randomized clinical trials using the log-rank test and univariate Cox analyses to produce hazard ratios with 95% CI. Multivariate Cox regression can be applied to identify independent factors associated with TTD.

The log-rank test is a non-parametric test to compare the survival distributions of two samples A and B. The null hypothesis H_0 is that the survival distribution in both groups A and B are equal, i.e., the expected probability of the event at time i is the same in both groups. Under this null hypothesis, the theoretical probability of the event at time i is:

$$\mathsf{P}_i = (E_{Ai} + E_{Bi})/(V_{Ai} + V_{Bi}) \tag{6}$$

where:

- E_{Ai} is the number of observed events in group A at time i;

- E_{Bi} is the number of observed events in group B at time i;

- V_{Ai} is the number of patients in group A who do not present the event at time i;

- V_{Bi} is the number of patients in group B who do not present the event at time i.

	HRQOL scores					TTD as compared to			TUDD as compared to								
									Baseline			Best previous			Immediately preceding		
Id	T_1	T_2	T_3	T_4	T_5	Baseline score	Best previous score	Immediately preceding score	1	2	3	1	2	3	1	2	3
1	61	59	55			T_3	T_3	T_3	T_3	T_3	T_3	T_3	T_3	T_3	T_3	T_3	T_3
2	61	61	56	51	59	T_3	T_3	T_3	T_3			T_3			T_3		
3	70	75	73	69	65	T_5	T_4		T_5	T_5	T_5	T_4	T_4	T_4			
4	51		58	50	55		T_4	T_4				T_4		T_4	T_4		T_4
5	64	61	58	47	59	T_3	T_3	T_4	T_3	T_3	T_3	T_3	T_3	T_3	T_4		
6	69	72		66			T_4	T_4				T_4	T_4	T_4	T_4	T_4	T_4
7	64	60	57	58	65	T_3	T_3		T_3			T_3					
8		61		53	61		T_4	T_4				T_4		T_4			
9	60	62	56	61	57	T_3	T_3					T_3	T_5	T_3	T_3		
10	55																

Table 2: HRQOL scores obtained at 5 timepoints T_1 to T_5 for 10 patients and time events according to each definition of TTD and TUDD (1: definition 1; 2: definition 2; 3: definition 3) with a 5-point MCID.

Then the log-rank statistic equals:

$$\chi^2_{exp} = \frac{(\sum_i E_{Ai} - \sum_i(\mathsf{P}_i \times V_{Ai}))^2}{\sum_i(\mathsf{P}_i \times V_{Ai})} + \frac{(\sum_i E_{Bi} - \sum_i(\mathsf{P}_i \times V_{Bi}))^2}{\sum_i(\mathsf{P}_i \times V_{Bi})} \qquad (7)$$

Under the null hypothesis, χ^2_{exp} is distributed according to a χ^2 distribution with one degree of freedom.

The Cox regression model link the instantaneous risk of event λ at time t with other covariates $X_1, \ldots, X_n$ as follows:

$$\lambda(t, X_1, \ldots, X_n) = \lambda_0(t) \times \exp(\sum_{i=1}^{n} \beta_i X_i) \qquad (8)$$

where $\lambda_0(t)$ corresponds to a basic risk and corresponds to an instantaneous risk of event at time t when all covariates are equal to 0.

sectionIllustrations of the TTD and TUDD definitions Event definition and censoring rules depend on the definition of deterioration used, as illustrated in both Figure 1 and Table 1. Table 2 summarizes several situations for patients:

- the variable *id* is the patient's identification number;

- T_1 to T_5 correspond to five HRQOL assessments;

- a deterioration corresponds to a score decrease.

There are some missing scores for several measures, illustrating the issue of intermittent or monotone missing data in HRQOL studies.

Table 2 illustrates different scenarios according to the event definition. This table summarizes, for each definition of TTD and TUDD, the time of the events for patients experiencing a deterioration with a 5-point MCID. In other cases, patients are censored at the time of the

	TTD as compared to the baseline score		Sensitivity analysis	
Id	Event	Time	Event	Time
1	1	$T_3 - T_1$	1	$T_3 - T_1$
2	1	$T_3 - T_1$	1	$T_3 - T_1$
3	1	$T_5 - T_1$	1	$T_5 - T_1$
4	0	$T_5 - T_1$	0	$T_5 - T_1$
5	1	$T_3 - T_1$	1	$T_3 - T_1$
6	0	$T_4 - T_1$	0	$T_4 - T_1$
7	0	$T_5 - T_1$	0	$T_5 - T_1$
8	0	0	1	0
9	0	$T_5 - T_1$	0	$T_5 - T_1$
10	0	1	1	1

Table 3: Time to deterioration compared to the baseline score with a 5-point MCID for the patients of Table 2 and sensitivity analysis considering patients with no baseline score or with no follow up score as events.

last available HRQOL measure. For example, for the TTD as compared to the baseline score, patients ♯4 and ♯6 do not present a deterioration and are censored at T_5 and T_4 respectively. Patient ♯2 presents a deterioration as compared to the baseline score at T_3, but not definitive as compared to the score qualifying the deterioration (TUDD.3). Patient ♯9 presents a deterioration as compared to the best score at T_3: at T_3, the previous best score equals 62 and the difference between this score and the score at T_3 is 6, i.e., greater than the MCID. However, the HRQOL level of the patient goes up to 61 and then the deterioration is not definitive as compared to the deteriorated score 56. At the last HRQOL assessment, HRQOL score equals 57 and the previous best score is still 62. In this way, the deterioration observed at T_3 is definitive as compared to the deteriorated score. We recall that if a deterioration is followed by missing data (patient dropped out), the deterioration is definitive, whatever the definition of TUDD retained.

Two variables have then been created:

- a dummy variable event indicating if the patient has deteriorated (event = 1) or not (event = 0);

- a time variable equals to the time between baseline date, and the date of deterioration or censoring.

Table 3 below illustrates these two variables for the 10 patients of Table 2 and one definition of time to deterioration.

To illustrate, the patient ♯1 is in deterioration as compared to the baseline score (event = 1) and the time between the baseline date and the date of the deterioration equals $T_3 - T_1$, then time = $T_3 - T_1$.

Patient ♯8 has no baseline score:

- in the primary analysis, this patient is censored at baseline (event = 0);

- in the sensitivity analysis, this patient is in deterioration at baseline (event = 1).

In both cases, the time to deterioration equals: time $= T_1 - T_1 = 0$.

In the same way, patient $\sharp 10$ has only a baseline score, and no follow-up score:

- in the primary analysis, this patient is censored one day after baseline (event $= 0$);

- in the sensitivity analysis, this patient is in deterioration one day after baseline (event $= 1$).

In both cases, the time to deterioration equals to: time $= T_1 + 1 - T_1 = 1$.

3. QoLR package

The **QoLR** package was developed to allow the longitudinal analysis of HRQOL according to the time to deterioration approach. The **QoLR** has dependencies with two R packages (**survival**, Therneau 2017, and **zoo**, Zeileis and Grothendieck 2005).

3.1. Package structure

The **QoLR** package contains a set of functions to calculate the scores of the EORTC QLQ-C30 and most of its modules and two other programs to determine the time to deterioration in HRQOL score as a modality of longitudinal analysis whatever the definition of deterioration used. Other programs make it possible to print all of the results or to output in a CSV file according to treatment arm and perform all sensitivity analyses according to one reference score. A last program was created to obtain the TTD curves calculated according to the Kaplan-Meier estimation method, with the option of displaying some information (number of patients at risk, cumulative number of events, hazard ratio and log-rank test if two treatment arms are compared).

For the convenience of the reader, we summarize all the main functions, with arguments and descriptions of our package **QoLR** in Table 4. Hereafter, we describe some of the package functions in detail.

Function `scoring.QLQC30()`

The first application of the **QoLR** package is the estimation of the scores of most of the EORTC HRQOL questionnaires, such as the QLQ-C30 cancer specific questionnaire.

The first argument of the function `scoring.QLQC30` and other functions for scoring is the name of the dataset with the items comprising the answers to the questionnaire (`X` parameter). The patient's identification number can be specified in the `id` parameter. A variable identifying the HRQOL assessment number can also be specified in the `time` parameter in case of a longitudinal HRQOL assessment.

The items must be named `q1` to `qi` for the QLQ-C30 ($i = 30$), QLQ-C15-PAL ($i = 15$) and IN-PATSAT32 ($i = 32$) questionnaire. For all other supplementary modules, items must be named `q31` to `qi`, because these modules have to be administered in conjunction with the QLQ-C30 core questionnaire. For example, items must be named `q1` to `q30` for the QLQ-C30 and then `q31` to `q50` for the QLQ-BN20. Moreover, the order of items has to be respected in the dataset.

Functions	Arguments	Description
scoring.QLQC30	(data, id, time)	Scoring of the EORTC QLQ-C30 questionnaire
scoring.QLQC15PAL	(data, id, time)	Scoring of the EORTC QLQ-C15-PAL questionnaire for palliative care
scoring.QLQBN20	(data, id, time)	Scoring of the EORTC QLQ-BN20 module
scoring.QLQBR23	(data, id, time)	Scoring of the EORTC QLQ-BR23 module
scoring.QLQCR29	(data, id, time)	Scoring of the EORTC QLQ-CR29 module
scoring.QLQCX24	(data, id, time)	Scoring of the EORTC QLQ-CX24 module
scoring.QLQELD14	data, id, time)	Scoring of the EORTC QLQ-ELD14 module for elderly cancer patients
scoring.QLQEN24	(data, id, time)	Scoring of the EORTC QLQ-EN24 module
scoring.QLQHCC18	(data, id, time)	Scoring of the EORTC QLQ-HCC18 hepatocellular carcinoma module
scoring.QLQHN35	(data, id, time)	Scoring of the EORTC QLQ-H&N35 module
scoring.QLQINFO25	(data, id, time)	Scoring of the EORTC QLQ-INFO25 information module
scoring.QLQLC13	(data, id, time)	Scoring of the EORTC QLQ-LC13 module
scoring.QLQLMC21	(data, id, time)	Scoring of the EORTC QLQ-LMC21 liver metastases in colorectal cancer module
scoring.QLQMY20	(data, id, time)	Scoring of the EORTC QLQ-MY20 module
scoring.QLQOES18	(data, id, time)	Scoring of the EORTC QLQ-OES18 module
scoring.QLQOG25	(data, id, items)	Scoring of the EORTC QLQ-OG25 module
scoring.QLQOV28	(data, id, time)	Scoring of the EORTC QLQ-OV28 ovarian module
scoring.QLQPR25	(data, id, time)	Scoring of the EORTC QLQ-PR25 module
scoring.QLQSTO22	(data, id, time)	Scoring of the EORTC QLQ-STO22 module
scoring.INPATSAT32	(data, id, time)	Scoring of the EORTC IN-PATSAT32 satisfaction with care questionnaire
TTD	(X, score, ...)	Estimation of the time to deterioration
TUDD	(X, score, ...)	Estimation of the time until definitive deterioration
plotTTD	(time, event, ...)	Kaplan-Meier curve of the TTD or TUDD
write.TTD	(X, score, ...)	Estimation of the TTD and print the results in a CSV file
write.TUDD	(X, score, ...)	Estimation of the TUDD and print the results in a CSV file

Table 4: Summary of the functions in the **QoLR** package.

The result is a data frame: each variable corresponds to a HRQOL score. The score names are the same as those used in the EORTC scoring manual (Fayers *et al.* 1999). If a patient's identification number was specified in the `id` parameter, then this variable is replicated in the data frame obtained. Moreover, if a variable identifying the HRQOL assessment number was specified in the `time` parameter, then this variable is also replicated in the data frame obtained.

Function `TTD()`

This function is used to estimate the time to deterioration of a HRQOL score. To apply this function, the dataset must respect a general structure. The dataset `X` must be in long format

with the following variables in this order:

1. Patient's identification number;

2. Variable identifying the HRQOL assessment number;

3. Date of HRQOL measure;

4. HRQOL scores;

5. Other variables like the date of death or the treatment arm.

The dataset must also be sorted by patient identification number and HRQOL measurement time. Dates must be in Julian format (i.e., number of days since a reference time point).

All definitions of TTD presented in this paper are programmed. Table 5 summarizes the arguments of this function and their possible values. According to the definition of deterioration retained, you must specify the following:

- The name of your dataset (`X`);

- The name of the HRQOL scores studied (`score = " "`);

- The value of the MCID (`MCID = 5`, for example);

- The reference score to qualify the deterioration which can be one of the following:

  - The baseline score (`ref.init = "baseline"`);
  - The best previous score (`ref.init = "best"`);
  - The immediately preceding score (`ref.init = "previous"`).

- Whether the deterioration corresponds to a decrease (`order = 1`) or an increase of the score (`order = 2`);

- Whether patients with no baseline score are excluded (`no_baseline = "excluded"`), censored (`no_baseline = "censored"`) or are in deterioration (`no_baseline = "event"`) since baseline;

- Whether patients with no follow-up score are censored (`no_follow = "censored"`) or are in deterioration (`no_follow = "event"`) one day after baseline;

- If death is considered as an event, you must specify the name of the variable in your dataset `X` which contains the date of death for patients who died during the study and with missing values for patients still alive at the end of the study.

An option (`sensitivity = TRUE`) makes it possible to perform all sensitivity analyses available in one application of the `TTD()` function.

For example, if you fixed `no_baseline = "censored"`, `no_follow = "censored"` and `sensitivity = TRUE` then:

- A first analysis is conducted censoring patients with no baseline and those with no follow-up score;

Arguments	Values
`X`	Matrix or data frame
`score`	Vector
`MCID`	Scalar
`ref.init`	= `"baseline"` (default)/= `"best"` /= `"previous"`
`order`	= 1 (default) /= 2
`no_baseline`	= `"censored"` (default)/= `"event"` /= `"excluded"`
`no_follow`	= `"censored"` (default)/= `"event"`
`death`	= `NA` (default) or vector
`sensitivity`	= `FALSE` (default)/= `TRUE`

Table 5: Arguments of the `TTD` function.

- Sensitivity analysis ♯1: A sensitivity analysis is conducted considering these patients in deterioration.

If, in addition to these parameters, the variable `death` is equals to the variable corresponding to the date of death in your dataset, then four analyses are performed:

- A first analysis censoring patients with no baseline, those with no follow-up and those who died without experiencing deterioration before dying;

- Sensitivity analysis ♯1: considering patients with no baseline and those with no follow-up score in deterioration;

- Sensitivity analysis ♯2: considering death as an event;

- Sensitivity analysis ♯3: considering simultaneously patients with no baseline, those with no follow-up score and death as an event.

The result of this function is a data frame with:

- the patient's identification number;

- a dummy variable called `event` equal to 1 if the patient is deteriorated, 0 if the patient is censored;

- a variable called `time`, equal to the time to censoring or the time to deterioration <u>in months</u>.

Since both variables `event` and `time` are created for each score treated and each definition of TTD, we added the name of the corresponding score as a suffix. For example, if `score = c("score1", "score2")`. Then four variables are created: `event.score1`, `time.score1`, `event.score2` and `time.score2`.

Moreover, if `sensitivity == TRUE`, then added variables `event` and `time` are created:

- `event.SA1` for sensitivity analysis ♯1: `event` and `event.SA1` are equals except for patients with no baseline score and those with no follow-up score, `event.SA1 = 1` while `event = 0`. Since `time.SA1 == time`, then `time.SA1` was omitted;

Arguments	Values
X	Matrix or data frame with the data
score	Vector
MCID	Vector
ref.init	= "baseline" (default)/= "best" /= "previous"
ref.def	= "def1"/= "def2"/ = "def3"
order	= 1 (default) /= 2
no_baseline	= "censored" (default)/= "event" /= "excluded"
no_follow	= "censored" (default)/= "event"
death	= NA (default) or vector
sensitivity	= FALSE (default)/= TRUE

Table 6: Arguments of the TUDD function.

- event.SA2 and time.SA2 for sensitivity analysis ♯2;

- event.SA3 for sensitivity analysis ♯3. Since time.SA3 == time.SA2, then time.SA3 was omitted.

Function TUDD()

This function allows the estimation of the time until definitive deterioration according to the definition retained. All definitions of TUDD presented in this paper are implemented. The syntax of this function is nearly the same as for TTD. Table 6 below summarizes the arguments of this function. Only one supplementary parameter is available compared to the TTD() function, namely: the parameter ref.def in which you can specify the notion of "definitive deterioration" according to the three definitions proposed in Section 2:

- With no further improvement of at least k points as compared to the reference score (ref.def = "def1");

- maintaining this deterioration of at least k points for all following scores, i.e., the deterioration is observed for all the following scores (ref.def = "def2");

- with no further improvement of at least k points as compared to the score qualifying the deterioration (ref.def = "def3").

Moreover, in this function, you can perform sensitivity analysis according to the MCID, thus the MCID parameter is a vector, not scalar.

Function plotTTD()

The **QoLR** package also contains a program called plotTTD() to obtain the TTD curves calculated according to the Kaplan-Meier estimation method for all patients or by treatment arm (only two groups are allowed). The time parameter is a vector equal to the time to deterioration or the time to censoring, and the event parameter is a dummy vector equal to 1 if the patient is deteriorated and 0 if not.

Other information can also be added using options, e.g., at regular time point t for all patients or by treatment arm:

Arguments	Values
time	Vector
event	Dummy vector
group	= NULL (default)/ vector
nrisk	= TRUE (default)/= FALSE
nevent	= FALSE (default)/= TRUE
group.names	= NULL/ vector
t	Vector
info	= FALSE (default)/= TRUE
pos.info	= NULL/ vector
xlab	= character
ylab	= character

Table 7: Arguments of the `plotTTD` function.

- number of patients at risk (`nrisk = T`);

- cumulative number of events (`nevents = T`).

In the case of TTD curves by treatment arm you must give the name of the group variable in the `group` parameter and the label of each group as you would like it to print in the `group.names` parameter. The hazard ratio with 95% confidence interval and log-rank test can also be added on the graph (`info = TRUE`) at a determined position specified by the user (`pos.info = c()`).

`xlab` and `ylab` correspond to the name of the horizontal and vertical axis respectively. Table 7 summarizes the arguments of this function.

This function makes it possible to plot the TTD curves with additional information useful for researchers to easily obtain standard curves for presentation or scientific publications.

Functions `write.TTD()` *and* `write.TUDD()`

These outputs can also be obtained for all definitions of TTD or TUDD with `write.TTD()` and `write.TUDD()` respectively, in the **QoLR** package. These programs create a comma-separated values (CSV) file with the results of the TTD or TUDD analyses performed in one or more scores according to one main deterioration definition. All sensitivity analyses according to this primary definition are also performed, with one or more MCID, for all patients or by group (e.g., treatment arm effect). The results produced are as follows:

- the number of patients initially at risk and the number of events;

- the median time to deterioration with 95% confidence interval;

and in the case of analyses performed by group (only two groups are allowed):

- the results of the log-rank test;

- the univariate hazard ratio with 95% confidence interval.

The arguments of this function are almost the same as the `TTD()` and `TUDD()` functions. An additional parameter corresponds to the name of the file in which to print the results (`file = ""`). The user can also specify the directory for this file in the (`file = ""`) parameter.

3.2. Applications of the QoLR package

To apply one or more of these functions, you need to load the **QoLR** library in the R current session with the following command:

```
R> library("QoLR")
```

We shall demonstrate the use of five main programs of **QoLR** in one dataset `dataqol` corresponding to a randomized clinical trial with the answers to the 30 items of the QLQ-C30 questionnaire for 40 patients with longitudinal HRQOL assessment. Patients in this table were randomly allocated to one treatment group corresponding to the variable `Arm` (dichotomous variable equal to 0 or 1).

This dataset is available in **QoLR** package and can be imported into R via the command `data("dataqol")`.

This dataset contains the following information:

- the patient identification number in the `Id` parameter;

- the treatment group to which each patient was allocated, corresponding to the variable `Arm` (dichotomous variable equals to 0 or 1);

- a variable indicating the theoretical HRQOL assessment number in the `time` parameter (equal to 0 for baseline measure);

- the date of HRQOL assessment in the `date` parameter;

- and the answers to the 30 items of the QLQ-C30 in the variables `q1` to `q30`;

- and the date of death for patients who died during the study in the `death` parameter (missing for all patients who were still alive at the end of the study).

This dataset is in long format. The variable `date` corresponds to the date of HRQOL measure in Julian format. The baseline date was set to 0 and all the following dates correspond to the time between the measure and the baseline date (the baseline date is the date of origin). HRQOL was evaluated about every 50 days. In the same way, `death` corresponds to the time between the baseline date and the date of death if the patient died during the study. Moreover, some items and/or measurement times are missing in order to illustrate how missing data are treated in the scoring and then in the Time to deterioration analysis.

This dataset looks like the following:

```
R> data("dataqol")
R> head(dataqol)
```

```
  Id Arm time date q1 q2 q3 q4 q5 q6 q7 q8 q9 q10 q11 q12 q13 q14 q15 q16 q17
1  1   2    0    0  1  2 NA  2  1  2  1  1  2   2   3   2   1   3   3   2   2
```

```
2  1  2    1   43  3  4  3  3  1  2  2  2  1   4   2   4   3   3   3   3   3
3  1  2    2   92  2  2  1  2  1  2  2  1  2   2   2   3   3   2   2   2   2
4  2  2    0    0  1  1  1  1  1  1  1  1  2   3   4   3   4   1   1   4   1
5  2  2    1   55  2  3  2  2  1  1  1  1 NA   3   3   3   3   1   1   4   1
6  2  2    2  149  2  3  2  2  2  1  1  2  2   3   3   4   3   1   1   4   1
   q18 q19 q20 q21 q22 q23 q24 q25 q26 q27 q28 q29 q30 death
1   2   2   1   2   4   2   3   1   2   2   1   5   4    NA
2   4   3   2   3   3   3   3   1   1   4   1   3   3    NA
3   2   2   2   3   3   2   3   1   2   3   2   4   4    NA
4   3   2   1   2   3   3   3   1   1   2   1   4   4   271
5   3   1   2   1   3   3   2   1   1   3   1   4   5   271
6   3   2   2   2   3   3   2   1   1   3   1   4   4   271
```

Scoring EORTC questionnaires

We illustrate the use of the program `scoring.QLQC30()` to estimate the score of a QLQ-C30 questionnaire in the `dataqol` dataset.

To apply the `scoring.QLQC30()` program, you have to specify at least the name of the dataset. You can also specify a patient's identification number in the `id` parameter as well as a time's identification number in the `time` parameter. The result is a data frame `score_dataqol` in which each variable corresponds to a score of the questionnaire, the first variable is the `id` parameter and the second one is the `time` parameter if they are filled in:

```
R> score_dataqol <- scoring.QLQC30(dataqol, id = "Id", time = "time")
R> head(round(score_dataqol))
```

```
  Id time QL  PF  RF EF  CF SF  FA NV PA DY  SL  AP  CO DI FI
1  1    0 58  83  83 42 100 67  33 67 33  0  67   0  33 33  0
2  1    1 33  40  67 33  83 50 100 67 33 33  33  67  67 67  0
3  1    2 50  80  67 42  83 50  44 33 33  0  33  67  33 33 33
4  2    0 50 100 100 42 100 83  67  0 33  0 100 100 100  0  0
5  2    1 58  67 100 58  83 67  67  0  0  0  67  67 100  0  0
6  2    2 50  60 100 50  83 67  78  0 33 33  67  67 100  0  0
```

Then, we can illustrate the `TTD` and `TUDD` functions.

Time to HRQOL score deterioration

Time to deterioration. To apply the `TTD` function, we have to first modify the `scoring_dataqol` dataframe in order to respect the requested format of the `TTD` function.

As a reminder, this dataset must contain the following information in the following order:

1. Patient identification number;

2. Variable identifying the HRQOL assessment number with value 0 for the baseline measure;

3. Date of HRQOL measure;

4. HRQOL scores;

5. Other variables, such as the date of death or the treatment arm.

The date of HRQOL assessment as well as the treatment arm and date of death are available in the `dataqol` dataset. We thus merged the `score_dataqol` dataframe with the important variables of the `dataqol` dataframe as follows:

```
R> info <- dataqol[, c("Id", "time","date", "death", "Arm")]
R> dataqol_final <- merge(score_dataqol, info, by = c("Id", "time"))
```

```
  Id time QL PF RF EF  CF SF  FA NV PA DY SL  AP  CO  DI FI date death Arm
1  1    0 58 83 83 42 100 67  33 67 33  0 67   0  33  33  0    0    NA   2
2  1    1 33 40 67 33  83 50 100 67 33 33 33  67  67  67  0   43    NA   2
3  1    2 50 80 67 42  83 50  44 33 33  0 33  67  33  33 33   92    NA   2
4 10    0 25 50 67 42   0 17 100 50 67  0  0 100 100   0 67    0    NA   1
5 10    1 17 33 67 25  67 17 100 67 83  0 67 100   0 100 67   56    NA   1
6 11    3 67 80 83 75 100 67  33 17 33  0  0  33   0   0 33  150   273   2
```

Then we reorganized the obtained `dataqol_final` dataset in order that the `date` variable appeared in third position:

```
R> dataqol_final <- dataqol_final[, c(1:2, 18, 3:17, 19:20)]
R> head(round(dataqol_final))
```

```
  Id time date QL PF RF EF  CF SF  FA NV PA DY SL  AP  CO  DI FI death Arm
1  1    0    0 58 83 83 42 100 67  33 67 33  0 67   0  33  33  0    NA   2
2  1    1   43 33 40 67 33  83 50 100 67 33 33 33  67  67  67  0    NA   2
3  1    2   92 50 80 67 42  83 50  44 33 33  0 33  67  33  33 33    NA   2
4 10    0    0 25 50 67 42   0 17 100 50 67  0  0 100 100   0 67    NA   1
5 10    1   56 17 33 67 25  67 17 100 67 83  0 67 100   0 100 67    NA   1
6 11    3  150 67 80 83 75 100 67  33 17 33  0  0  33   0   0 33   273   2
```

Finally, in order to apply the `TTD` function, the dataset has to be sorted by patient's identification number (`Id`) and HRQOL measures (`time`):

```
R> dataqol_final <- dataqol_final[order(dataqol_final$time), ]
R> dataqol_final <- dataqol_final[order(dataqol_final$Id), ]
R> head(round(dataqol_final))
```

```
   Id time date QL  PF  RF EF  CF SF  FA NV PA DY  SL  AP  CO DI FI death Arm
1   1    0    0 58  83  83 42 100 67  33 67 33  0  67   0  33 33  0    NA   2
2   1    1   43 33  40  67 33  83 50 100 67 33 33  33  67  67 67  0    NA   2
3   1    2   92 50  80  67 42  83 50  44 33 33  0  33  67  33 33 33    NA   2
25  2    0    0 50 100 100 42 100 83  67  0 33  0 100 100 100  0  0   271   2
26  2    1   55 58  67 100 58  83 67  67  0  0  0  67  67 100  0  0   271   2
27  2    2  149 50  60 100 50  83 67  78  0 33 33  67  67 100  0  0   271   2
```

Since HRQOL was measured at several measurement times for each patient, we can study the time to deterioration of the HRQOL scores.

As a reminder, for global health status (variable `QL`) and other functional scales (variables `PF` for physical functioning to `SF` for social functioning) a high score reflects a high level of HRQOL or functioning. For symptomatic scales (variables `FA` for fatigue to `FI` for financial difficulties), a high score reflects a high symptomatic level.

To begin, we can study the reference definition of TTD of the `QL` score, i.e., a deterioration as compared to the baseline score (`ref.init = "baseline"`) with at least 5-point MCID (`MCID = 5`) and considering patients with no baseline or with no follow-up measure censored at baseline or just after baseline (`no_baseline = "censored"` and `no_follow = "censored"`). The score `QL` corresponds to a measure of global HRQOL. In this way, a deterioration corresponds to a decrease of the score (`order = 1`). Since the values of these parameters are the default values (except for the MCID), we do not need to specify their values and the function can thus be applied as follows:

```
R> ttd1 <- TTD(dataqol_final, score = "QL", MCID = 5)
R> head(ttd1)

  Id event.QL    time.QL
1  1        1 1.41273101
2  2        0 4.89527721
3  3        1 1.77412731
4  4        0 0.03285421
5  5        0 0.00000000
6  6        1 1.70841889
```

The result is a data frame with the identification number of patients (`Id`), the time to deterioration or to censoring in months (`time.QL`) and a dummy variable (`event.QL`) indicating whether the patient is deteriorated (`event.QL = 1`) or not (`event.QL = 0`). The suffix "QL" corresponds to the name of the treated score. According to this definition, 24 patients are deteriorated among the 40 patients in terms of the `QL` score:

```
R> sum(ttd1$event.QL, na.rm = TRUE)
```

```
[1] 24
```

If we want to consider patients with no baseline or no follow up as events, we have to fix the parameters `no_baseline` and `no_follow` to `"event"` as follows:

```
R> ttd2=TTD(dataqol_final, score = "QL", order = 1, MCID = 5,
+    no_baseline = "event", no_follow = "event")
R> head(ttd2)

  Id event.QL    time.QL
1  1        1 1.41273101
2  2        0 4.89527721
```

```
3   3         1 1.77412731
4   4         1 0.03285421
5   5         1 0.00000000
6   6         1 1.70841889

R> sum(ttd2$event.QL, na.rm = TRUE)

[1] 33
```

9 patients with no baseline or no follow-up measure are then added to the events.

To consider death as an event, we have to specify the value of the `death` parameter:

```
R> ttd3 <- TTD(dataqol_final, score = "QL", MCID = 5, death = "death")
R> head(ttd3)

  Id event.QL  time.QL
1  1        1 1.412731
2  2        1 8.903491
3  3        1 1.774127
4  4        1 1.018480
5  5        0 0.000000
6  6        1 1.708419
```

Finally, we can obtain directly all the sensitivity analyses along with the primary analysis in one application of the function `TTD` by specifying `sensitivity = TRUE`:

```
R> ttd4 <- TTD(dataqol_final, score = "QL", MCID = 5, death = "death",
+    sensitivity = TRUE)
R> head(ttd4)

  Id event.QL    time.QL event.SA1.QL event.SA2.QL time.SA2.QL event.SA3.QL
1  1        1 1.41273101            1            1    1.412731            1
2  2        0 4.89527721            0            1    8.903491            1
3  3        1 1.77412731            1            1    1.774127            1
4  4        0 0.03285421            1            1    1.018480            1
5  5        0 0.00000000            1            0    0.000000            1
6  6        1 1.70841889            1            1    1.708419            1
```

When all sensitivity analyses are performed, some added variables are created. Variables `event.QL` and `time.QL` still correspond to the results of the primary analysis (TTD as compared to the baseline score in the present case). The variable `event.SA1.QL` is the dummy event variable considering patients with no baseline or no follow up measure in deterioration since baseline, whereas they were censured in the primary analysis (Sensitivity Analysis ♯1). Since the corresponding times are the same as in the primary analysis, no new time variable was created. Variables `event.SA2.QL` and `time.SA2.QL` are the results for the analysis adding death as an event (Sensitivity Analysis ♯2). Finally, `event.SA3.QL` corresponds to the event variable with death and patients with no baseline or no follow up as events (Sensitivity Analysis ♯3).

In order to integrate the response shift effect, we can choose the best previous HRQOL score or the immediately preceding score as the reference score by specifying `ref.init = "best"` or `ref.init = "previous"` respectively.

In the following example, the reference score is the best previous `QL` score:

```
R> ttd5 <- TTD(dataqol_final, score = "QL", MCID = 5, ref.init = "best")
R> head(ttd5)

  Id event.QL    time.QL
1  1        1 1.41273101
2  2        1 4.89527721
3  3        1 1.77412731
4  4        0 0.03285421
5  5        0 1.34702259
6  6        1 1.70841889

R> sum(ttd5$event.QL, na.rm = TRUE)

[1] 26
```

In this case, 26 patients experienced a deterioration as compared to the best previous `QL` score

The function `TTD()` can handle many scores simultaneously, functional and/or symptomatic scores. You must define the name of the scores studied in the `score` parameter as well as the order to consider (i.e., decrease or increase): `order = 1` for global health status or functional scales and `order = 2` for symptomatic scales. Variables `event` and `time` are then created for each score with the score name as a suffix. The following example represents the application of the `TTD` function as compared to the baseline score with a 5-point MCID for `QL`, `PF` (with `order = 1` for both scores) and `FA` (with `order = 2`):

```
R> ttd6 <- TTD(dataqol_final, score = c("QL", "PF", "FA"), order = c(1, 1, 2),
+     MCID = 5)
R> head(ttd6)

  Id event.QL    time.QL event.PF    time.PF event.FA    time.FA
1  1        1 1.41273101        1 1.41273101        1 1.41273101
2  2        0 4.89527721        1 1.80698152        1 4.89527721
3  3        1 1.77412731        0 5.15811088        0 5.15811088
4  4        0 0.03285421        0 0.03285421        0 0.03285421
5  5        0 0.00000000        0 1.34702259        0 0.00000000
6  6        1 1.70841889        1 1.70841889        1 1.70841889
```

Time until definitive deterioration. The TUDD is studied with the `TUDD()` function, quite similar to the `TTD()` function. By default, the deterioration is defined as a deterioration with a k-point MCID as compared with the baseline score, with no further improvement of more than k points as compared to the baseline score (Bonnetain *et al.* 2010). The result of the application of this function is fairly similar to that of the `TTD()`. However, for `TUDD()`, the value of the MCID is also specified in the variable names `time` and `event`.

The result of the reference definition of TUDD is as follows:

```
R> tudd1 <- TUDD(dataqol_final, score = "QL", MCID = 5)
R> head(tudd1)

  Id event.5.QL   time.5.QL
1  1         1   1.41273101
2  2         0   4.89527721
3  3         1   1.77412731
4  4         0   0.03285421
5  5         0   0.00000000
6  6         0  11.49897331

R> sum(tudd1$event.5.QL, na.rm = TRUE)

[1] 22
```

Thus, 22 patients experienced a definitive deterioration as compared to the baseline score according to the definition of Bonnetain *et al.* (2010).

The deterioration can also be definitive as compared to the deterioration observed, i.e., with no further improvement of 5-point MCID as compared to the score obtained at the time of the first deterioration. This definition is applied by setting the parameter `ref.def` to the value `"def3"`:

```
R> tudd2 <- TUDD(dataqol_final, score = "QL", MCID = 5, ref.def = "def3")
R> head(tudd2)

  Id event.5.QL   time.5.QL
1  1         1   3.02258727
2  2         0   4.89527721
3  3         0   5.15811088
4  4         0   0.03285421
5  5         0   0.00000000
6  6         0  11.49897331
```

As for the TTD, all sensitivity analyses can be performed simultaneously with the primary definition of TUDD. Moreover, many MCID can be specified. In fact, as defined in section 2, we need a dependence between sensitivity analyses varying the MCID for TUDD. An indicator of the MCID value is added as a suffix of the resulting parameters `event` and `time`. The following example is based on the physical functioning score (`PF`):

```
R> tudd3 <- TUDD(dataqol_final, score = "PF", MCID = c(5,10),
+    sensitivity = TRUE)
R> head(round(tudd3, 2))

  Id event.10.PF time.10.PF event.10.SA1.PF event.5.PF time.5.PF event.5.SA1.PF
1  1          1        1.41               1          1       1.41              1
2  2          1        1.81               1          1       1.81              1
3  3          0        5.16               0          0       5.16              0
4  4          0        0.03               1          0       0.03              1
5  5          0        1.35               0          0       1.35              0
6  6          0       11.50               0          1       1.71              1
```
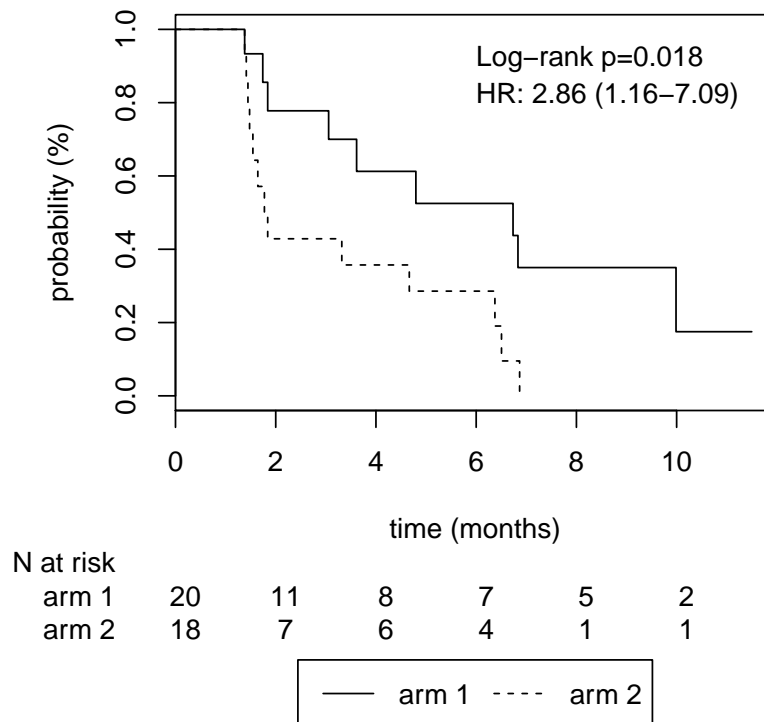
Figure 2: Kaplan-Meier survival curve for the time until definitive `QL` score deterioration of at least 5-point MCID as compared to the baseline score.

In this application, death has not been taken into account. Only two sensitivity analyses were performed: the first regarding patients with no baseline or no follow-up measure and the second on the MCID value.

Time to deterioration curves. Figure 2 corresponds to the TUDD of `QL` score as compared to the baseline score with a 5-point MCID according to treatment arm (`Arm` parameter). In this graph, we printed the number of patients still at risk at each time point according to treatment arm (`nrisk = T`). Moreover, the result of the log-rank test and the hazard ratio of Arm 2 vs. Arm 1 is also printed (`info = T, pos.info = c(6, 0.8)`). The hazard ratio (Arm 2 vs. Arm 1) equals 2.86 with 95% confidence interval $(1.16 - 7.09)$ and the result of the log-rank test is $p = 0.018$.

```
R> tudd1 <- TUDD(dataqol_final, score = "QL", MCID = 5, ref.init = "baseline",
+    ref.def = "def1")
R> ttd_1 <- merge(tudd1, unique(dataqol_final[, c("Id", "Arm")]))
R> plotTTD(ttd_1$time.5.QL, ttd_1$event.5.QL, ttd_1$Arm, nrisk = T, nevent = F,
+    group.names = c("Arm 1", "Arm 2"), t = seq(0, 10, 2), info = T,
+    pos.info = c(6, 0.8), xlab = "time (months)", ylab = "probability (%)")
```

Write all the results of a time to deterioration analysis in a CSV file. All information regarding TTD or TUDD analysis can also be output to a CSV file such as the number

	MCID = 5 points TTD baseline N (events)	Median (CI 95%)	Log-rank	HR (CI 95%)
`QL`				
Arm 1	20 (11)	3.61 (1.84–NA)	$p = 0.136$	1
Arm 2	18 (13)	1.81 (1.54–NA)		1.87 (0.81–4.31)
`PF`				
Arm 1	20 (10)	1.84 (1.71–NA)	$p = 0.634$	1
Arm 2	20 (11)	1.82 (1.54–NA)		1.23 (0.52–2.93)
`FA`				
Arm 1	20 (8)	6.74 (1.74–NA)	$p = 0.35$	1
Arm 2	20 (11)	4.67 (1.64–NA)		1.55 (0.61–3.92)

Table 8: The CSV file created with the application of the `write.TTD` function.

of events and the effect size of the treatment effect with hazard ratio and log-rank test with `write.TTD()` or `write.TUDD()` commands available in **QoLR** package.

For example, the following command creates a CSV file named `"file_TTD_baseline"` which is located in the current directory of your R session. You can also specify the directory in the `file` parameter. With this command, we studied the TTD of `QL`, `PF` and `FA` scores with a 5-point and a 10-point MCID as compared to the baseline score.

Table 8 is an extract of this file for a 5-point MCID.

```
R> write.TTD(dataqol_final, score = c("QL", "PF", "FA"), order = c(1, 1, 2),
+    MCID = c(5, 10), group = "Arm", names.group = c("Arm 1", "Arm 2"),
+    sensitivity = FALSE, file = "file_TTD_baseline")
```

For example, in arm 1 and arm 2, 11 and 13 patients respectively experienced a deterioration of at least 5 points in the `QL` score as compared to the baseline score, among the 38 patients included (20 patients in arm 1, 18 patients in arm 2, no data were available for two patients). The median TTD was 3.61 months with a 95% confidence interval (1.81–NA) for arm 1 and 1.81 months with a 95% confidence interval (1.54–NA) for arm 2 (log-rank $p = 0.136$). The univariate Cox hazard ratio of arm 2 vs. arm 1 was 1.87 with a 95% confidence interval (0.81–4.31).

4. Conclusion and outlook

The **QoLR** package is the first R package dedicated to the analysis of HRQOL. The implementation of the time to deterioration definitions in a HRQOL score allows the dissemination of these approaches in order to achieve the goal of standardization of longitudinal HRQOL analysis in oncology clinical trials.

QoLR will be updated as new modules are developed by the EORTC HRQOL group. The package will be completed over time, by some simulation algorithms of longitudinal HRQOL data with intermittent or monotone missing data of type Missing Completely At Random or Missing Not At Random for example. Other programs will make it possible to print the results of univariate and multivariate Cox analyses in a CSV file. Moreover, an ongoing

project is investigating the presence of competitive risk between events. The current package will then be expanded by the addition of competitive risk models. Finally, the time to deterioration approach can be implemented using SAS software (SAS Institute Inc. 2011) to allow researchers not familiar with R software to apply this longitudinal analysis method.

Acknowledgments

References

Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JC, others (1993). "The European Organization for Research and Treatment of Cancer QLQ-C30: A Quality-of-Life Instrument for Use in International Clinical Trials in Oncology." *Journal of the National Cancer Institute*, **85**(5), 365–376. `doi:10.1093/jnci/85.5.365`.

Ahmed S, Mayo NE, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R (2005). "Change in Quality of Life of People with Stroke Over Time: True Change or Response Shift?" *Quality of Life Research*, **14**(3), 611–627. `doi:10.1007/s11136-004-3708-0`.

Anota A, Hamidou Z, Paget-Bailly S, Chibaudel B, Bascoul-Mollevi C, Auquier P, Westeel V, Fiteni F, Borg C, Bonnetain F (2013). "Time to Health-Related Quality of Life Score Deterioration as a Modality of Longitudinal Analysis for Health-Related Quality of Life Studies in Oncology: Do We Need RECIST for Quality of Life to Achieve Standardization?" *Quality of Life Research*, pp. 1–14. `doi:10.1007/s11136-013-0583-6`.

Bonnetain F, Dahan L, Maillard E, Ychou M, Mitry E, Hammel P, Legoux JL, Rougier P, Bedenne L, Seitz JF (2010). "Time Until Definitive Quality of Life Score Deterioration as a Means of Longitudinal Analysis for Treatment Trials in Patients with Metastatic Pancreatic Adenocarcinoma." *European Journal of Cancer*, **46**(15), 2753–2762. `doi:10.1016/j.ejca.2010.07.023`.

Bullinger M (2002). "Assessing Health Related Quality of Life in Medicine. An Overview Over Concepts, Methods and Applications in International Research." *Restorative Neurology and Neuroscience*, **20**(3), 93–101. `doi:10.1097/00004356-200209000-00005`.

Cnaan A, Laird NM, Slasor P (1997). "Tutorial in Biostatistics: Using the General Linear Mixed Model to Analyse Unbalanced Repeated Measures and Longitudinal Data." *Statistics in Medicine*, **16**, 2349–2380. `doi:10.1002/(sici)1097-0258(19971030)16:20<2349::aid-sim667>3.0.co;2-e`.

Cocks K, King MT, Velikova G, St-James MM, Fayers PM, Brown JM (2011). "Evidence-Based Guidelines for Determination of Sample Size and Interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30." *Journal of Clinical Oncology*, **29**(1), 89–96. `doi:10.1200/jco.2010.28.0107`.

Diggle PJ (1988). "An Approach to the Analysis of Repeated Measurements." *Biometrics*, pp. 959–971. `doi:10.2307/2531727`.

Fairclough DL (2010). *Design and Analysis of Quality of Life Studies in Clinical Trials*. 2nd edition. Chapman & Hall/CRC. `doi:10.1201/9781420061185`.

Fayers PM, Aaronson NK, Bjordal K, Curran D, Grønvold M (1999). "EORTC QLQ-C30 Scoring Manual." *Technical report*, EORTC.

Gibbons FX (1999). "Social Comparison as a Mediator of Response Shift." *Social Science & Medicine*, **48**(11), 1517–1530. `doi:10.1016/s0277-9536(99)00046-5`.

Goel MK, Khanna P, Kishore J (2010). "Understanding Survival Analysis: Kaplan-Meier Estimate." *International Journal of Ayurveda Research*, **1**(4), 274. `doi:10.4103/0974-7788.76794`.

Hamidou Z, Dabakuyo TS, Mercier M, Fraisse J, Causeret S, Tixier H, Padeano MM, Loustalot C, Cuisenier J, Sauzedde JM, others (2011). "Time to Deterioration in Quality of Life Score as a Modality of Longitudinal Analysis in Patients with Breast Cancer." *The Oncologist*, **16**(10), 1458–1468. `doi:10.1634/theoncologist.2011-0085`.

Hunger M, Döring A, Holle R (2012). "Longitudinal Beta Regression Models for Analyzing Health-Related Quality of Life Scores Over Time." *BMC Medical Research Methodology*, **12**(1), 144. `doi:10.1186/1471-2288-12-144`.

Jaeschke R, Singer J, Guyatt GH (1989). "Measurement of Health Status: Ascertaining the Minimal Clinically Important Difference." *Controlled Clinical Trials*, **10**(4), 407–415. `doi:10.1016/0197-2456(89)90005-6`.

Korfage IJ, de Koning HJ, Essink-Bot ML (2007). "Response Shift Due to Diagnosis and Primary Treatment of Localized Prostate Cancer: A Then-Test and a Vignette Study." *Quality of Life Research*, **16**(10), 1627–1634. `doi:10.1007/s11136-007-9265-6`.

Lydick E, Epstein R (1993). "Interpretation of Quality of Life Changes." *Quality of Life Research*, **2**(3), 221–226. `doi:10.1007/bf00435226`.

Norman GR, Sloan JA, Wyrwich KW (2003). "Interpretation of Changes in Health-Related Quality of Life: The Remarkable Universality of Half a Standard Deviation." *Medical Care*, **41**(5), 582–592. `doi:10.1097/01.mlr.0000062554.74615.4c`.

Osoba D (2011). "Health-Related Quality of Life and Cancer Clinical Trials." *Therapeutic Advances in Medical Oncology*, **3**(2), 57–71. `doi:10.1177/1758834010395342`.

Osoba D, Rodrigues G, Myles J, Zee B, Pater J (1998). "Interpreting the Significance of Changes in Health-Related Quality-of-Life Scores." *Journal of Clinical Oncology*, **16**(1), 139–144. `doi:10.1159/000227647`.

Pan AW, Chen YL, Chung LI, Wang JD, Chen TJ, Hsiung PC (2012). "A Longitudinal Study of the Predictors of Quality of Life in Patients with Major Depressive Disorder Utilizing a Linear Mixed Effect Model." *Psychiatry Research*, **198**(3), 412–419. `doi:10.1016/j.psychres.2012.02.001`.

Pauler DK, McCoy S, Moinpour C (2003). "Pattern Mixture Models for Longitudinal Quality of Life Studies in Advanced Stage Disease." *Statistics in Medicine*, **22**(5), 795–809. doi:10.1002/sim.1397.

Penar-Zadarko B, Binkowska-Bury M, Wolan M, Gawelko J, Urbanski K (2012). "Longitudinal Assessment of Quality of Life in Ovarian Cancer Patients." *European Journal of Oncology Nursing*, **17**(3), 139–5. doi:10.1016/j.ejon.2012.10.014.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

SAS Institute Inc (2011). *The SAS System, Version 9.3*. Cary. URL http://www.sas.com/.

Sprangers MAG, Schwartz CE (1999). "Integrating Response Shift into Health-Related Quality of Life Research: A Theoretical Model." *Social Science & Medicine*, **48**(11), 1507–1515. doi:10.1016/s0277-9536(99)00045-3.

Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, Croft P, de Vet HCW (2010). "Mind the MIC: Large Variation Among Populations and Methods." *Journal of Clinical Epidemiology*, **63**(5), 524–534. doi:10.1016/j.jclinepi.2009.08.010.

Therneau TM (2017). *A Package for Survival Analysis in S*. R package version 2.41-2, URL https://CRAN.R-project.org/package=survival.

Ubel PA, Peeters Y, Smith D (2010). "Abandoning the Language of "Response Shift": A Plea for Conceptual Clarity in Distinguishing Scale Recalibration from True Changes in Quality of Life." *Quality of Life Research*, **19**(4), 465–471. doi:10.1007/s11136-010-9592-x.

Wiklund I (2004). "Assessment of Patient-Reported Outcomes in Clinical Trials: The Example of Health-Related Quality of Life." *Fundamental & Clinical Pharmacology*, **18**(3), 351–363. doi:10.1111/j.1472-8206.2004.00234.x.

Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD (1999a). "Linking Clinical Relevance and Statistical Significance in Evaluating Intra-Individual Changes in Health-Related Quality of Life." *Medical Care*, **37**(5), 469–478. doi:10.1097/00005650-199905000-00006.

Wyrwich KW, Tierney WM, Wolinsky FD (1999b). "Further Evidence Supporting an SEM-Based Criterion for Identifying Meaningful Intra-Individual Changes in Health-Related Quality of Life." *Journal of Clinical Epidemiology*, **52**(9), 861–873. doi:10.1016/s0895-4356(99)00071-2.

Zeileis A, Grothendieck G (2005). "**zoo**: S3 Infrastructure for Regular and Irregular Time Series." *Journal of Statistical Software*, **14**(6), 1–27. doi:10.18637/jss.v014.i06.

Affiliation:

Amélie Anota, Franck Bonnetain
Quality of Life in Oncology Clinical Research Platform
Methodology and Quality of Life in Oncology Unit (EA 3181)
University Hospital of Besançon
25 030 Besançon Cedex, France
E-mail: aanota@chu-besancon.fr, franck.bonnetain@univ-fcomte.fr

Marion Savina
INSERM
Clinical and Epidemiological Research Unit (CIC-EC 7) – CTD INCa
Institut Bergonie
33 000 Bordeaux, France
E-mail: Marion.Savina@isped.u-bordeaux2.fr

Caroline Bascoul-Mollevi
Biostatistics Unit
Cancer Institute
34 090 Montpellier Val d'Aurelle, France
E-mail: Caroline.Mollevi@icm.unicancer.fr