Available online at www.sciencedirect.com

ScienceDirect

journal homepage: http://www.elsevier.com/locate/rpor

Original research article

Why we should take care of the competing risk bias in survival analysis: A phase II trial on the toxicity profile of radiotherapy for prostate cancer

Annarita Tullio[a,*], *Alessandro Magli*[b], *Eugenia Moretti*[c], *Francesca Valent*[a]

[a] *Hygiene and Clinical Epidemiology Unit, "S. Maria della Misericordia" University Hospital of Udine, Udine, Italy*
[b] *Department of Radiation Oncology, "S. Maria della Misericordia" University Hospital of Udine, Udine, Italy*
[c] *Department of Medical Physics, "S. Maria della Misericordia" University Hospital of Udine, Udine, Italy*

ARTICLE INFO

Article history:
Received 30 October 2018
Received in revised form
21 May 2019
Accepted 11 August 2019
Available online 19 August 2019

Keywords:
Competing risks
Cumulative incidence function
Fine and Gray
Subdistribution hazard
Survival analysis

ABSTRACT

Aim: The aim of the present study is to evaluate and quantify the bias of competing risks in an Italian oncologic cohort comparing results from different statistical analysis methods.

Background: Competing risks are very common in randomized clinical trials and observational studies, in particular oncology and radiotherapy ones, and their inappropriate management causes results distortions widely present in clinical scientific articles.

Materials and methods: This is a single-institution phase II trial including 41 patients affected by prostate cancer and undergoing radiotherapy (IMRT-SIB) at the University Hospital of Udine.

Different outcomes were considered: late toxicities, relapse, death.

Death in the absence of relapse or late toxicity was considered as a competing event.

Results: The Kaplan Meier method, compared to cumulative incidence function method, overestimated the probability of the event of interest (toxicity and biochemical relapse) and of the competing event (death without toxicity/relapse) by 9.36%. The log-rank test, compared to Gray's test, overestimated the probability of the event of interest by 5.26%.

The Hazard Ratio's and cause specific hazard's Cox regression are not directly comparable to subdistribution hazard's Fine and Gray's modified Cox regression; nonetheless, the FG model, the best choice for prognostic studies with competing risks, found significant associations not emerging with Cox regression.

Conclusions: This study confirms that using inappropriate statistical methods produces a 10% overestimation in results, as described in the literature, and highlights the importance of taking into account the competing risks bias.

* *Corresponding author.*
E-mail address: annarita.tullio@uniud.it (A. Tullio).

https://doi.org/10.1016/j.rpor.2019.08.001

1. Background

Time-to-event outcomes are very common in randomized clinical trials and observational studies.

In these cases, a competing risk is an event whose occurrence precludes the observation of the primary event of interest.[1–4]

Competing risks are a well-known problem but scientific literature, including high impact journals, is widely bias affected.

Three articles, published in 2010,[5] in 2015[6] and 2017,[1] describe such bias in the recent medical literature.

The first study[5] reports that 37 papers, of the 50 considered, published between 2008 and 2011, had at least one endpoint definition implying the presence of competing risks, and that only 2 articles had correctly used the curves based on the cumulative incidence function (CIF) in the presence of competing risks.

The second research[6] was performed on MEDLINE to identify all published studies in 2013 in which a survival analysis was performed using the Kaplan Meier (KM) method in the 'core clinical journals'. This restriction was deemed useful as it allowed to identify articles published in well-known journals easily accessible to clinicians.

Of the 100 articles identified, forty-six studies had conducted a KM analysis susceptible to the bias. Sixteen studies had provided the number of competing events, allowing to estimate the bias; of these, a third had overestimated by 10% or more the risk of outcome using the KM method.

Austin and Fine[1] conducted a review of randomized controlled trials with survival outcomes published in high-impact general medical journals.

Of the 40 studies identified, published in the last 3 months of 2015, 31 were potentially susceptible to competing risks and only 5 reported CIF.

Articles with survival outcomes susceptible to bias of competing events have proved to be significantly present in journals with a high impact factor (≥4.901). This is because the KM method is easier to understand and implement with common statistical packages.

Nonetheless, the KM method does not fit very well to predict the probability of a certain outcome in the presence of competing events. It can handle only one event at a time: all the other events are treated as censored observations and often competing risks are mistakenly conceptualized as censoring events[7].

The assumption of independent censoring is also violated, which means patients who experience a competing event will no longer have the same possibility of developing the event of interest as patients that continue to be included in the study. As a result, the KM method overestimates the probability of the event of interest giving misleading results in the presence of competing risks.

The alternative method to KM for competing risk management is based on the CIF and takes into account all the events[3,8] and the informative nature of the competing events censoring.[9]

Competing events are "censored" in a particular way remaining in the risk set[10] up to a hypothetical censoring.[4,9,11]

The cumulative incidence of an event assessed considering competing risks is the probability of experiencing the event of interest within a certain time and not experience a competing event within the same time.

In the presence of competing events, the CIF is estimated for the event of interest and for all the competing events and their estimates depend on one another.[3,8]

When there is only one type of event, the estimate of the cumulative incidence using the KM method and the competing risks method is the same. Similarly, in a study design, if we are interested in the cumulative incidence estimate only of the first event, in the presence of multiple events, there are no competing events.

In the literature there are many examples of how the KM method overestimates the probability of both the event of interest and the competing one, by 10%,[6] while the probability of not having events remains unchanged.[2]

The Kaplan Meier method is therefore an inadequate statistic in the presence of competing risks and the use of the CIF method is recommended.

The cumulative incidence in various groups can be compared using non-parametric tests such as the log-rank test when censoring is independent and incidences are based on the KM method, or Gray's test,[12] in the presence of competing risks, by comparing the underlying subdistribution hazards.

In the absence of censoring for competing events, Gray's test and the log-rank test yield identical results.

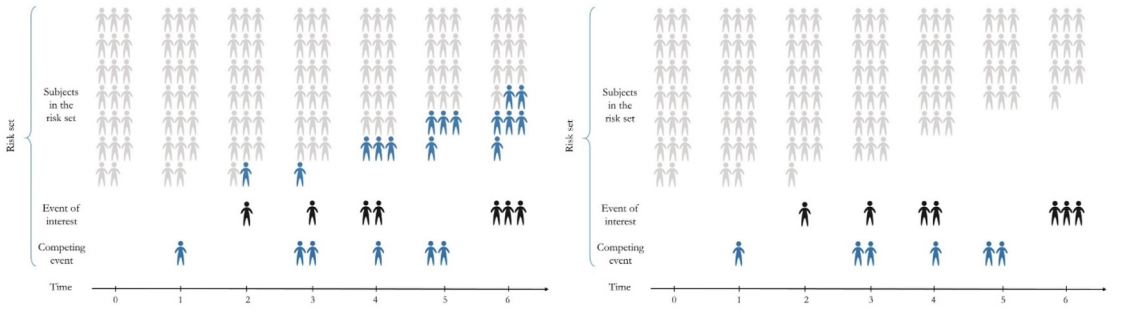
In a competing risk setting, to identify the correct regression method, we need to know whether the research is etiological or prognostic.[2] We can choose between two types of a hazard function: the cause specific hazard (CSH), representing the instantaneous rate of a cause, and the subdistribution hazard (SDH), representing the hazard corresponding to the cumulative incidence function for a cause.[13]

Etiological studies use CSH to estimate the effect size, while prognostic studies use the model of risk subdistribution, a modified Cox regression model proposed by Fine and Gray[14] in 1999 defining the hazard of the cumulative incidence function. The difference between the two types of hazard function lies in the risk set. In the cause specific hazard, the risk set decreases at any time when an event occurs for a cause other than the event of interest, while in the subdistribution hazard the subjects that experience a competing event remain in the risk set (rather than being censored and eliminated), although, in reality, they are no longer at risk of experiencing the event of interest, up to a potential censoring, as clarified in Fig. 1.

For etiological research, the CSH model may be more appropriate than the SDH method because the regression parameters estimated with this method directly quantify the HR among the subjects that are actually at risk of developing the event of interest. In this case, Cox regression analysis is applied for each of the specific events.

Differently, an SDH represents a ratio in a non-existent population, including those who have experienced the competing event and could no longer experience the event of interest. This measure is useful for the prediction and prognosis, in order to evaluate realistic clinical situations.[2]

Right now all the main statistical packages, like SAS,[15–17] STATA[18] and R (i.e. 'cmprsk',[19] 'gcerisk',[20] 'survminer'[21] packages), are able to manage relatively friendly competing risks



Hazard	t=0	t=1	t=2	t=3	t=4	t=5	t=6
SDH	0/20=0	0/20=0	1/20=0.5	1/19=0.05	2/18=0.11	0/16=0	3/16=0.19

Hazard	t=0	t=1	t=2	t=3	t=4	t=5	t=6
CSH	0/20=0	0/20=0	1/19=0.5	1/18=0.06	2/15=0.13	0/12=0	3/10=0.33

Fig. 1 – Different risk set for different regression methods: SDH vs CSH.Adapted from Noordzij et al.[2]

bias. It is basic to define carefully the outcome and to apply the correct censoring method before managing survival analyses.

The objective of our study was to evaluate and quantify the bias of competing risks comparing results from different censoring and statistical analysis methods.

2. Methods

Data used in this study come from a single-institution phase II trial including 41 patients affected by prostate cancer and undergoing radiotherapy (IMRT-SIB) from 2009 to 2012 at the University Hospital of Udine (ASUIUD).[22]

The research ethics board of the University Hospital of Udine (ASUIUD) approved the study, and all the patients enrolled signed an informed consent, according to the Helsinki Declaration.

Eligible patients had histologically confirmed prostate cancer with at least one of the following high-risk features: clinical stage T2c/T3 or PSA level ≥20 but <100 ng/mL or Gleason score ≥8.[22,23]

Exclusion criteria comprised lymph nodes involvement, distant metastasis, a history of inflammatory bowel disease, active collagen vascular disease, previous malignancy within 5 years of prostate carcinoma diagnosis (except non-melanomatous skin cancer), prior pelvic RT, hip prosthesis and severe urinary symptoms.[22]

IMRT-SIB treatment planning is described elsewhere.[22]

Once a week during treatment full blood work, physical examination and toxicity evaluation were performed.

Following the end of radiotherapy, patients were monitored on a quarterly basis, in the first 2 years, and on a 6-month basis, in the next 3 years and then annually. Follow-up period continued until last observation or patients' death.

Different outcomes were considered: late rectal toxicity, late gastrointestinal toxicity, late genitourinary toxicity, relapse, death.

Primary end points related to late toxicity were defined as any event persisting or occurring more than 3 months after RT completion and biochemical relapse within 6 years. The toxicity events were prospectively recorded using the Radiation Therapy Oncology Group (RTOG) morbidity criteria.[23,24]

According to the Phoenix definition, biochemical recurrence was defined, as an increase in the PSA level of 2 ng/mL or more above the PSA nadir after RT.[25]

Death in the absence of relapse or late toxicity was considered as a competing event.

Each type of event was analyzed separately.

The end of the follow up was set at 31/03/2017.

3. Statistical analyses

The Simon's two-stage design method was conducted to determine the sample size.[26]

The trial was planned to evaluate primarily rectal toxicity, considering as acceptable a late toxicity rate ≤5%, and considering as unacceptable a late toxicity rate ≥25%, with a 90% power and a significance level of 5%. The stopping rule laid down by the investigation was ≥4 patients with 41 patients enrolled.

The study population features were investigated performing descriptive statistics on categorical and numerical variables. Frequency distributions were used for categorical variables. For numerical variables, we considered mean, median, interquartile range, standard deviation, 25° and 75° percentile, minimum and maximum values. Kolmogorov–Smirnov test was performed for checking normality of the distribution for numerical variables.

A study on toxicity profiles was conducted with two survival analyses: one survival analyses without considering competing risks, i.e., censoring patient who died and patients who completed follow up without any toxicity (or relapse) with no distinction; one survival analysis applying an informative censoring to patients who died without any toxicity or relapse, i.e., taking competing events into account. Both death and toxicity (or relapse) events can be considered as competing events one for the other.

In the analyses conducted without competing events, we used the Kaplan Meier estimator (survival estimate), log-rank test (survival estimate, *p*-value), Cox regression univariate and age-adjusted models (HR, *p*-value).

The survival analyses with competing risks were managed through the calculation of cumulative incidence function (CIF), the homogeneity test of Gray (CIF, *p*-value), Cox regression model obtaining the cause specific hazard (CSH, *p*-value)

and Fine and Gray modified Cox regression model with the subdistribution hazard (SDH, p-value), both univariate and age adjusted.

In both cases we produced survival tables and curves.

Results of KM and CIF estimates, as well as the results from log-rank test and Gray's test, were compared through the percentage difference, considering as 100% reference CIF values and then considering proportionally 1-KM values. The overestimation was calculated as the difference between the complementary to 1 of the KM estimate and CIF value.

Before performing regression analyses, the Cox regression proportionality assumption was checked using the Schoenfeld's residuals; only the variables respecting the assumption ($p \geq 0.05$) were included in the models.

Age adjusted analyses were conducted considering two age-groups: <or ≥ median age at the beginning of radiotherapy (72 years).

Dosimetric–toxicity relation was investigated with unadjusted and age adjusted analyses considering dosimetric parameters, such as the mean and median dose and dose distribution parameters, such as V45, V50, V66, V70 (percentage irradiated tissue for dose) evaluated as numeric continuous variables.

Since the results of the CSHs model and Cox regression (HRs) were the same, they were reported just once as CSHs.

We compared CSHs and SDHs and calculated their percentage differences, considering as 100% reference SDH values and then considering proportionally CSH values.

All statistical analyses were performed using SAS© software, version 9.4 (SAS institute, Inc., Cary, NC, USA) and R 3.4.2. The significance level was set at 0.05.

4. Results

Patient characteristics are summarized in Table 1. Median age was 72 years (range, 51–81 years), 12.2% of patients had diabetes, 24.4% were on anti-coagulant therapy and 53.7% suffered from hypertension. Median follow-up duration was between 62.39 and 62.85 months (range, 0.00–92.68 months).

The CIF and 1-KM estimates of event (toxicity or relapse) and competing event (death without toxicity or relapse) were compared and the results shown in Figs. 2 and 3 and Table 2.

Probability estimates curves of events and competing events using CIF and 1-KM methods are illustrated in Fig. 2. Percentage probabilities of events, competing events and censoring at the end of follow up and their sum with KM and CIF estimates are illustrated in the stacked barplot in Fig. 3, considering as 100% reference the sum of the three with the CIF method.

Table 2 presents CIF and 1-KM estimates with their percentage difference, considering CIF values as 100% reference and considering 1-KM values proportionally. The KM method overestimates the probability of the event and the competing event by 9.36% (mean value; range: 1.58–19.61) considering the events separately.

Results of survival analyses stratified by groups through CIF and 1-KM methods comparing the groups with log-rank test and Gray's test are shown in Fig. 4 and Table 3. The KM method overestimates the probability of event by 5.26% (median value), considering Gray's test results as 100% reference and considering those of the log-rank test proportionally.

There are no differences in significance between log-rank and Gray's test.

The results of Cox regression model (CSHs) and Fine and Gray's modified Cox regression model (SDHs) with their differences, both univariate and age adjusted, are shown in Table 4.

Table 1 – Patients characteristics.

Characteristic	Value
Number of patients	41
Age, yr	
Mean ± SD	71.83 ± 6.14
Median	72.00
Range (min–max)	51.00–81.00
IQR	6.00
Late rectal toxicity event, N (%)	5 (12.20)
Late gastrointestinal toxicity event, N (%)	7 (17.07)
Late genitourinary toxicity event, N (%)	6 (14.63)
Relapse, N (%)	3 (7.32)
Death, N (%)	8 (19.51)
Follow-up LRT, mo	
Median	62.39
Range (min–max)	4.24–92.68
IQR	21.62
Follow-up LGiT, mo	
Median	62.85
Range (min–max)	0.00–92.68
IQR	35.09
Follow-up LGuT, mo	
Median	62.39
Range (min–max)	5.98–92.68
IQR	22.80
Follow-up relapse, mo	
Median	62.85
Range (min–max)	6.05–92.68
IQR	21.42
With diabetes, N (%)	5 (12.20)
With hypertension, N (%)	22 (53.66)
With oral anti-coagulant, N (%)	10 (24.40)
Rectal dosimetric parameters	
Mean* ± SD/median° ± IQR (min–max)	
Mean dose (Gy)	35.58* ± 2.06 (33.25–42.62)
Median dose (Gy)	39.03* ± 1.27 (36.17–41.55)
V 50 (%)	16.12* ± 2.43 (11.44–20.66)
V 66 (%)	1.92* ± 1.58 (0.01–5.80)
V 70 (%)	0.04° ± 0.31 (0.00–2.69)
Bladder dosimetric parameters	
Mean* ± SD/median° ± IQR (min–max)	
Mean dose (Gy)	45.30* ± 2.72 (40.57–53.63)
Median dose (Gy)	43.38* ± 3.19 (37.72–53.49)
V 50 (%)	27.67° ± 11.76 (11.74–68.08)
V 60 (%)	9.35* ± 4.35 (2.28–19.38)
V 65 (%)	4.48* ± 2.58 (0.53–9.91)
V 70 (%)	0.36° ± 0.95 (0.00–4.32)

SD: standard deviation; IQR: interquartile range; Yr: years; mo: months; LRT: late rectal toxicity; LGiT: late gastrointestinal toxicity; LGuT: late genitourinary toxicity; N (%) number of patients with percentages in parentheses; *Mean ± SD; °Median ± IQR.

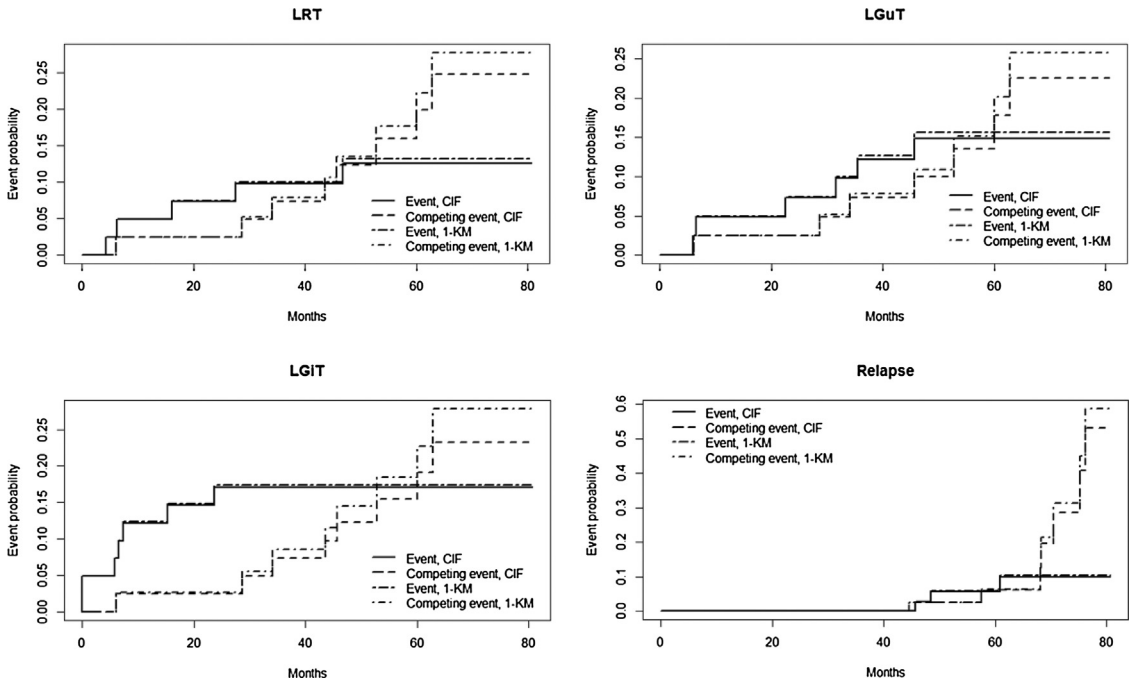


Fig. 2 – CIF and 1-KM curves for events and competing events. LRT: late rectal toxicity; LGiT: late gastrointestinal toxicity; LGuT: late genitourinary toxicity; 1-KM: complementary to 1 of the Kaplan Meier estimate; CIF: cumulative incidence function.

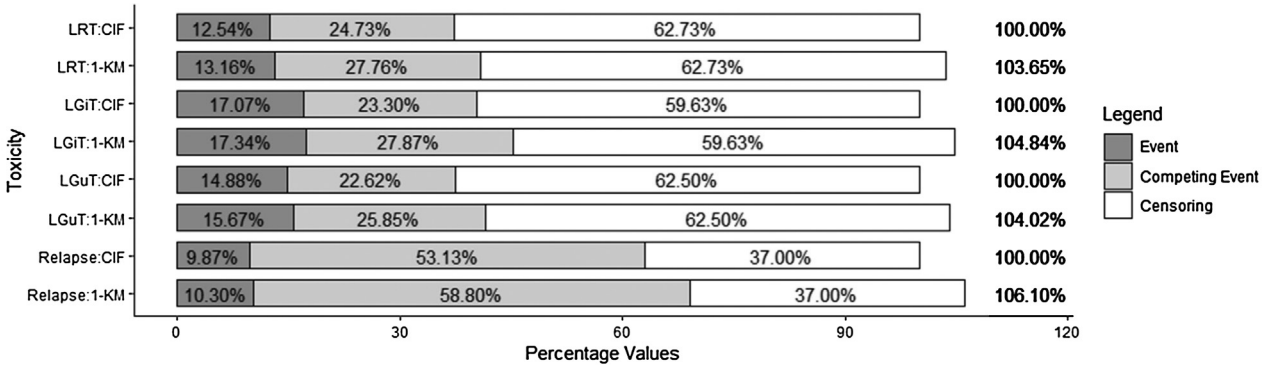


Fig. 3 – Percentage probabilities of toxicity/relapse, death without toxicity/relapse, censoring at the end of follow up without toxicity/relapse and their sum with KM and CIF estimates. LRT: late rectal toxicity, LGiT: late gastrointestinal toxicity; LGuT: late genitourinary toxicity; 1-KM: complementary to 1 of the Kaplan Meier estimate; CIF: cumulative incidence function.

Table 2 – The complementary to 1 of the KM estimate and the CIF estimate for the toxicity/relapse event and for the death event and their percentage difference ((1-KM)-CIF) considering CIF values as 100% reference and considering 1-KM values proportionally.

Outcome	Toxicity/relapse probability estimate					Death probability estimate				
	1-KM	CIF	E (*n*)	CE (*n*)	Diff. %	1-KM	CIF	E (*n*)	CE (*n*)	Diff. %
LRT	0.1316	0.1254	5	8	+4.94	0.2776	0.2473	8	5	+12.25
LGiT	0.1734	0.1707	7	8	+1.58	0.2787	0.233	8	7	+19.61
LGuT	0.1567	0.1488	6	7	+5.31	0.2585	0.2226	7	6	+16.13
Relapse	0.1030	0.0987	3	7	+4.36	0.5880	0.5313	7	3	+10.67

LRT: late rectal toxicity; LGiT: late gastrointestinal toxicity; LGuT: late genitourinary toxicity; 1-KM: complementary to 1 of the Kaplan Meier estimate; CIF: cumulative incidence function; E (*n*) number of events; CE (*n*) number of competing events.

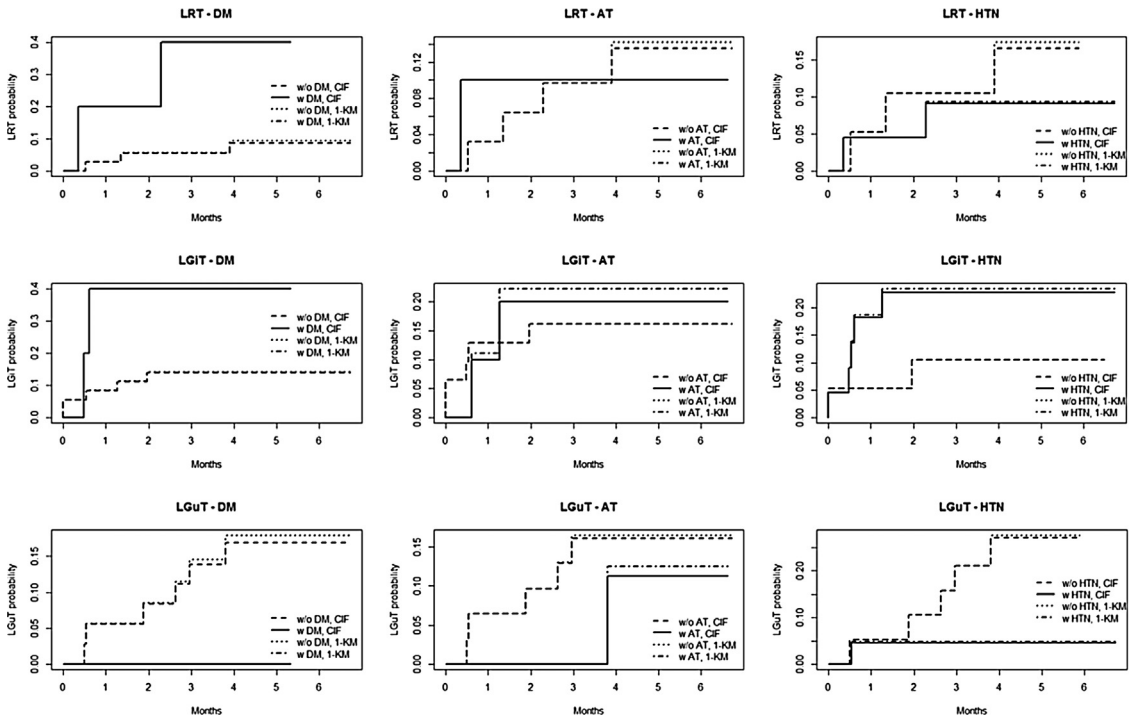


Fig. 4 – CIF and 1-KM curves for events and competing events stratified by diabetes mellitus, anticoagulation therapy, hypertension. LRT: late rectal toxicity; LGiT: late gastrointestinal toxicity; LGuT: late genitourinary toxicity; 1-KM: complementary to 1 of the Kaplan Meier estimate; CIF: cumulative incidence function; DM: diabetes mellitus; AT: anticoagulation therapy; HTN: hypertension.

Table 3 – Log-rank test and Gray'test *p*-values and cumulative incidence values and their percentage differences, considering Gray's test results as 100% reference and considering log-rank test ones proportionally.

Outcome	IV		Cumulative incidence		Diff. %	*p*-value	
			Log-rank test	Gray's test		Log-rank test	Gray's test
LRT	Diabetes	w	0.4000	0.4000	/	**0.04**	**0.04**
		w/o	0.0949	0.0877	+8.21		
	Anticoag.th.	w	0.1000	0.1000	/	0.90	0.85
		w/o	0.1419	0.1353	+4.88		
	Hypertension	w	0.0932	0.0910	+2.42	0.60	0.53
		w/o	0.1741	0.1657	+5.07		
LGiT	Diabetes	w	0.4000	0.4000	/	0.10	0.15
		w/o	0.1414	0.1389	+1.80		
	Anticoag.th.	w	0.2222	0.2000	+11.10	0.80	0.83
		w/o	0.1613	0.1613	/		
	Hypertension	w	0.2344	0.2273	+3.12	0.30	0.30
		w/o	0.1053	0.1053	/		
LGuT	Diabetes	w	/	/	/	0.30	0.34
		w/o	0.1792	0.1688	+6.16		
	Anticoag.th.	w	0.1250	0.1120	+11.61	0.60	0.60
		w/o	0.1650	0.1613	+2.29		
	Hypertension	w	0.0476	0.0455	+4.62	0.07	0.05
		w/o	0.2763	0.2713	+1.84		

LRT: late rectal toxicity; LGiT: late gastrointestinal toxicity; LGuT: late genitourinary toxicity; w: with; w/o: without.
bold values are <0.05.

The median value of overestimate by CSH is 0.68% (range −3.90 to +24.49) for univariate regression and 1.92% (range −4.67 to +25.64) for age adjusted regression, considering SDH results as 100% reference and considering CSH values proportionally.

5. Discussion

The aim of this study was to evaluate and quantify the bias of competing risks in an Italian oncologic cohort comparing the results obtained with methods considering

Table 4 – SDH and CSH values and their percentage difference for numerical continuous and dichotomous variables univariate and age adjusted regression, considering SDH results as 100% reference and considering CSH values proportionally.

Type	Outcome	IV	Univariate					Age adjusted				
			SDH	*p*	CSH	*p*	Diff. %	SDH	*p*	CSH	*p*	Diff. %
Dichotomous variables	LRT	Age	0.21	0.17	0.21	0.16	−1.44					
		Diabetes	5.89	0.04	5.66	0.06	−3.90	5.99	0.038	5.71	0.057	−4.67
		Hypertension	0.57	0.53	0.60	0.58	+5.08	0.64	0.60	0.69	0.68	+7.81
		Anticoag.th.	0.81	0.85	0.87	0.90	+7.16	1.04	0.97	1.29	0.82	+24.04
	LGiT	Age	1.10	0.90	1.06	0.94	−3.54					
		Diabetes	3.33	0.13	3.28	0.16	−1.62	3.33	0.14	3.28	0.16	−1.50
		Hypertension	2.31	0.31	2.39	0.30	+3.33	2.31	0.31	2.40	0.30	+3.90
		Anticoag.th.	1.20	0.82	1.28	0.77	+6.70	1.18	0.83	1.27	0.78	+7.63
	LGuT	Age	0.16	0.08	0.15	0.09	−1.29					
		Hypertension	0.16	0.10	0.18	0.11	+9.38	0.19	0.13	0.21	0.15	+10.53
		Anticoag.th.	0.57	0.58	0.60	0.64	+5.63	0.78	0.80	0.98	0.99	+25.64
Numerical continue variables	LRT	Mean dose	1.35	0.17	1.35	0.17	+0.15	1.37	0.21	1.37	0.19	0
		Median dose	1.30	0.52	1.30	0.48	+0.46	1.26	0.58	1.27	0.51	+0.79
		V 50	1.18	0.28	1.19	0.36	+1.11	1.20	0.28	1.21	0.37	+1.34
		V 66	1.44	0.02	1.58	0.11	+9.36	1.30	0.15	1.43	0.21	+10
		V 70	1.46	0.32	1.82	0.31	+24.49	1.16	0.74	1.43	0.53	+23.28
	LGuT	Mean dose	0.96	0.81	0.95	0.76	−0.42	0.94	0.72	0.93	0.62	−1.60
		Median dose	0.96	0.79	0.95	0.71	−0.84	0.93	0.71	0.90	0.53	−3.12
		V 50	1.00	0.95	1.01	0.89	+0.30	1.01	0.77	1.01	0.72	+0.30
		V 60	1.08	0.48	1.08	0.38	+0.56	1.06	0.52	1.07	0.43	+1.04
		V 65	1.17	0.33	1.18	0.29	+1.29	1.08	0.54	1.11	0.48	+2.31
		V 70	1.51	0.29	1.52	0.21	+0.79	1.25	0.54	1.27	0.44	+1.92

LRT: late rectal toxicity; LGiT: late gastrointestinal toxicity; LGuT: late genitourinary toxicity; CSH: cause specific hazard; SDH: subdistribution hazard.

competing events with those obtained with traditional ones.

The KM method overestimated the probability of events and competing events.

The KM method can deal just with one event at a time. Any other event has to be censored, the censoring is not informative and, in the presence of competing risks, the KM assumption of independent censoring is violated. The complementary to 1 of the KM estimation must be considered as the probability of the event of interest in a hypothetical world where a patient can fall only for the event of interest and no other case.

Therefore, methods for managing competing events must take into account the informative nature of censoring. The cumulative incidence method allows to manage both the event of interest and the competing event, which is then censored in an "informative" way.

Consequently, as seen from our results, in the presence of competing risks, KM, compared to CIF, overestimates the probability of the event of interest and of competing event, giving misleading results.

In this study we observed an overestimation of 9.36% (range 1.58–19.61), consistent with literature where the overestimation of the probability of the event of interest is reported to be around 10% or more.[6] The overestimation is the greater the more frequent and prior to the events of interest are the competing events. For example, the minimum overestimation (+1.58%) observed related to gastroenteric toxicity. There were 6 events and 8 competing events. The first event of interest occurred in the absence of competing events. The following 5 events of interest were concentrated between the 1st and the 2nd competing event (CE). In this case, the overestimation was caused by just 1 CE out of 8. On the other hand, the highest overestimation (+19.61%) observed was about death without LGiT. This was because deaths, i.e., the events of interest, took place after the occurrence of most of the LGiT events considered as CE.

Evaluating the percentage sum of the probabilities of the event of interest (LRT, LGiT, LGuT, relapse), of the competing event (death without toxicity or recurrence), and of censoring (survive without toxicity or recurrence), considering as 100% the reference the sum of the three (event of interest, competing event and censoring) with the CIF method, we observed an overestimation with the Kaplan Meier method which appears to be lower than the overestimation considered for the individual events. This is because patients censored who survived without events had the same probabilities with both methods and, in our study, they were more numerous than those who experienced an event, causing a "dilution" of the overestimation determined by competing events seen before.

Just as Kaplan Meier, the log-rank test, in the presence of competing events, determines an overestimate of the Gray test's results.

In some cases, we observed no overestimation from log-rank test estimates based on the KM method compared to Gray's test estimates based on CIF; this did not happen because there were no competing events in these strata, but because of the position of the competing event: the last event of interest happened before the first CE. In this way, competing events did not have the possibility to cause an increase of the estimate's probability of the event of interest.

Considering the example of the LRT in the presence of diabetes, we observed the last LRT event at 27 months and the first competing event at 63 months.

Considering the log-rank and Gray's test *p*-values we observed no differences. Both point out a statistically significant difference in LRT between patients with and without diabetes, a relation well known in literature. Many studies[27–30] indicate diabetes history as the first risk factor for developing radiotherapy-related late rectal toxicity events and for developing LRT events before non-diabetic patients[29] as seen in this study (4 months vs 6 months).

Regression analyses were performed using the Cox model and the model of Cox modified by Fine and Gray. The simple Cox regression can be used in the same way to estimate both the HR in the absence of competing events and the cause specific hazard (CSH) in the presence of competing risks.

Cox model modified by Fine and Gray estimates the subdistribution hazard (SDH) in the presence of competing events. However, it must be remembered that the SDH is not the same as the CSH and that different approaches to manage competing risks, as seen with our data, can give different results, explained by the different composition of the risks set. With regard to unadjusted analyses, CSH tends to give higher results than SDH in a greater number of cases. The percentage difference between the results obtained with the two techniques (CSH–SDH) ranges from −3.90% to +24.49%, with a median of +0.68%. With regard to age adjusted analyses (<72 years, ≥72), even in this case, the CSH tends to give higher results compared to SDH in a greater number of cases. The percentage difference among the results obtained with the two techniques (CSH–SDH) ranges from −4.67% to +25.64%, with a median of +1.92%. The differences observed between CSH and SDH depend, as mentioned, on the fact that SDH is actually a measure different from HR, representing a ratio in a non-existent population, including those who have experienced the competing event and could no longer experience the event of interest, whose interpretation is more difficult.

The HR of 5.66, e.g., related to the relationship between diabetes and rectal toxicity obtained with the causal approach, indicates that at any time after the first three months after the end of radiotherapy, diabetic patients had a risk of experiencing a rectal toxicity event 5.66 times higher than non-diabetic patients, among patients alive without any rectal toxicity event and still present in the risk set at that time.

The SDH greater than one (SDH = 5.89), obtained with the subdistribution model, means that the cumulative incidence of rectal toxicity is greater in diabetic patients than in non-diabetic patients.

However, the value of 5.89 is not easy to interpret as it expresses the rate of rectal toxicity (toxicity rate ratio) among patients who are alive without any rectal toxicity event or who died without any toxicity event at that time.

This measure is useful for prognostic research, as in our case, rather than to investigate etiological problems.

Regarding the significance of the relationship between the outcome and the predictors considered, using the SDH we can notice some well known[27–33] significances that are lost with the CSH in the unadjusted analysis, as the relation dose (V66) – percentage of irradiated tissue (rectum) ($p = 0.017$) and diabetes ($p = 0.042$) for rectal toxicity. The relevant relationship between diabetes and late rectal toxicity is also significant in the age adjusted regression analyses with the subdistribution method, while lost with the causal method ($p = 0.038$ vs $p = 0.057$).

The correct analyses method is therefore fundamental to study clinical data and to draw conclusions on important correlations.

Unfortunately, the existence of competing risks is yet a little known bias even if it has an important impact.

In 2017[1] Jason P. Fine, the father of the Fine and Gray modified Cox regression model for competing events management, has underlined the high prevalence of the problem in high-impact general medical journals and provided advice and recommendations for improvement.

It is of primary importance that clinicians know the existence of the competing risk bias, be able to recognize it and understand the implications of the results. It is crucial not only for those who write articles but mostly for those who read studies potentially susceptible.

To that end, it would be helpful if important research quality checklists, like CONSORT and STROBE, could implement an item about the presence/absence/management of competing risk bias.

6. Study limitations

The difference of about 10% that we found between the results of the classic methods of survival analyses and the methods of competing risks accounting is consistent with literature even if the low number of patients involved and events observed could have overestimated it.

It would be interesting to test different methods of survival analyses on a wider number of patients and events to better understand and quantify results and distortions caused by inconsistent analysis methods.

7. Conclusions

Competing risks are often present in survival analyses, especially in oncology and radiotherapy studies, and correct statistics should be applied to avoid distortions in the results. We suggest CIF and Gray's test should be used for survival analyses to avoid overestimations in the presence of competing risks and to account for competing risks in regression analyses using the Fine and Gray's modified Cox regression model for prognostic studies, not to miss potentially significant results.

Ethical statement

The research ethics board of the University Hospital of Udine (ASUIUD) approved the study, and all the patients enrolled signed an informed consensus, according to the Helsinky Declaration.

Conflict of interest

None declared.

Financial disclosure

None declared.

REFERENCES

1. Austin PC, Fine JP. Accounting for competing risks in randomized controlled trials: a review and recommendations for improvement. *Statist Med* 2017;**36**:1203–9.
2. Noordzij M, Leffondré K, Van Stralen KJ, et al. When do we need competing risks methods for survival analysis in nephrology? *Nephrol Dial Transplant* 2013;**28**:2670–7.
3. Scherzer R. *A tutorial on accounting for competing risks in survival analysis*. Clinical Research Statistical Methods Seminar, Kidney Health Research Collaborative, San Francisco VA Medical Center, University of California at San Francisco; 2017.
4. Hinchliffe SR. *Competing risks – what, why, when and how? Survival analysis for junior researchers*. Department of Health Sciences, University of Leicester; 2012.
5. Koller MT, Raatz H, Steyerbergc EW, et al. Competing risks and the clinical community: irrelevance or ignorance? *Stat Med* 2012;**31**:1089–97.
6. Van Walraven C, McAlistere FA. Competing risk bias was common in Kaplan–Meier risk estimates published in prominent medical journals. *J Clin Epidemiol* 2016;**69**, 170–173.e8.
7. Lambert PC, Dickman PW, Nelson CP, et al. Estimating the crude probability of death due to cancer and other causes using relative survival models. *Stat Med* 2010;**29**:885–95.
8. Dignam JJ, Kocherginsky MN. Choice and interpretation of statistical tests used when competing risks are present. *J Clin Oncol* 2008;**26**:4027–34.
9. Dignam JJ, Zhang Q, Kocherginsky MN. The use and interpretation of competing risks regression models. *Clin Cancer Res* 2012;**18**:2301–8.
10. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol* 2009;**170**:244–56.
11. Satagopan JM, Ben-Porat L, Berwick M, et al. A note on competing risks in survival data analysis. *J Cancer* 2004;**91**:1229–35.
12. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Ann Stat* 1988;**16**:1141–54.
13. Lee M, Ha ID, Lee Y. Frailty modeling for clustered competing risks data with missing cause of failure. *Stat Methods Med Res* 2017;**26**:356–73.
14. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc* 1999;**94**:496–509.
15. Lin G, So Y, Johnston G. *Analyzing survival data with competing risks using SAS® Software*. SAS Institute Inc.; 2012. Paper 344-2012, SAS Global Forum.
16. So Y, Lin G, Johnston G. *Using the PHREG procedure to analyze competing-risks data*. SAS Institute Inc.; 2014.
17. Gondara L. *Competing risk survival analysis using PHREG in SAS 9.4*. Cancer Surveillance & Outcomes (CSO), Population Oncology, BC Cancer Agency; 2015.
18. Gutierrez RG. *Competing-risks regression*. Statistics StataCorp LP, Stata Conference Boston; 2010.
19. Gray B. *cmprsk: subdistribution analysis of competing risks. R package version 2.2-7*; 2014.
20. Shen H, Carmona R, Mell L. *gcerisk: generalized competing event model. R package version 18.02.22*; 2018.
21. Kassambara A, Kosinski M. *survminer: drawing survival curves using 'ggplot2'. R package version 0.4.2*; 2018.
22. Magli A, Moretti E, Tullio A, et al. Hypofractionated simultaneous integrated boost (IMRT-SIB) with pelvic nodal irradiation and concurrent androgen deprivation therapy for high-risk prostate cancer: results of a prospective phase II trial. *Prostate Cancer Prostatic Dis* 2018;**21**:269–76.
23. Cooperberg MR, Pasta DJ, Elkin EP, et al. The University of California, San Francisco Cancer of the Prostate Risk Assessment score: a straightforward and reliable preoperative predictor of disease recurrence after radical prostatectomy. *J Urol* 2006;**175**:2369.
24. Cox JD, Stetz J, Pajak TF. Toxicity criteria of the Radiation Therapy Oncology Group (RTOG) and the European Organization for Research and Treatment of Cancer (EORTC). *Int J Radiat Oncol Biol Phys* 1995;**31**:1341–6.
25. Roach III M, Hanks G, Thames Jr H, et al. Defining biochemical failure following radiotherapy with or without hormonal therapy in men with clinically localized prostate cancer. Recommendations of the RTOG-ASTRO phoenix consensus conference. *Int J Radiat Oncol Biol Phys* 2006;**65**: 965–74.
26. Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials* 1989;**10**:1–10.
27. Maggio A, Magli A, Rancati T, et al. Daily sodium butyrate enema for the prevention of radiation proctitis in prostate cancer patients undergoing radical radiation therapy: results of a multicenter randomized placebo-controlled dose-finding phase 2 study. *Int J Radiat Oncol Biol Phys* 2014;**89**:518–24.
28. Akimoto T, Muramatsu H, Takahashi M, et al. Rectal bleeding after hypofractionated radiotherapy for prostate cancer: correlation between clinical and dosimetric parameters and the incidence of grade 2 or worse rectal bleeding. *Int J Radiat Oncol Biol Phys* 2004;**15**:1033–9.
29. Michalski JM, Gay H, Jackson A, et al. Radiation dose-volume effects in radiation-induced rectal injury. *Int J Radiat Oncol Biol Phys* 2010;**76**:S123–9.
30. Herold DM, Hanlon AL, Hanks GE. Diabetes mellitus: a predictor for late radiation morbidity. *Int J Radiat Oncol Biol Phys* 1999;**43**:475–9.
31. Barnett GC, De Meerleer G, Gulliford SL, et al. The impact of clinical factors on the development of late radiation toxicity: results from the Medical Research Council RT01 trial (ISRCTN47772397). *Clin Oncol (R Coll Radiol)* 2011;**23**:613–24.
32. Fuentes-Raspall R, Inorizac JM, Rosello-Serrano A, et al. Late rectal and bladder toxicity following radiation therapy for prostate cancer: predictive factors and treatment results. *Rep Pract Oncol Radiother* 2013;**18**: 298–303.
33. Fiorino C, Cozzarini C, Vavassori V, et al. Relationships between DVHs and late rectal bleeding after radiotherapy for prostate cancer: analysis of a large group of patients pooled from three institutions. *Radiother Oncol* 2002;**64**:1–12.