EPIDEMIOLOGY

Prediction of long-term cumulative incidences based on short-term parametric model for competing risks: application in early breast cancer

B. Cabarrou[1] · L. Belin[2] · S. M. Somda[1] · M. C. Falcou[3] · J. Y. Pierga[4,5] ·
Y. Kirova[6] · J. P. Delord[7] · B. Asselain[2] · T. Filleron[1]

Abstract Use of parametric statistical models can be a solution to reduce the follow-up period time required to estimate long-term survival. Mould and Boag were the first to use the lognormal model. Competing risks methodology seems more suitable when a particular event type is of interest than classical survival analysis. The objective was to evaluate the ability of the Jeong and Fine model to predict long-term cumulative incidence. Survival data recorded by Institut Curie (Paris) from 4761 breast cancer patients treated and followed between 1981 and 2013 were used. Long-term cumulative incidence rates predicted by the model using short-term follow-up data were compared to non-parametric estimation using complete follow-up data. 20- or 25-year cumulative incidence rates for loco-regional recurrence and distant metastasis predicted by the model using a maximum of 10 years of follow-up data had a maximum difference of around 6 % compared to non-parametric estimation. Prediction rates were underestimated for the third and composite event (contralateral or second cancer or death). Predictive ability of Jeong and Fine model on breast cancer data was generally good considering the short follow-up period time used for the estimation especially when a proportion of patient did not experience loco-regional recurrence or distant metastasis.

Keywords Breast cancer · Prediction · Parametric model · Competing risks

✉ T. Filleron
filleron.thomas@iuct-oncopole.fr

1   Biostatistic Unit, Bureau des Essais Cliniques, Institut Claudius Regaud, Institut Universitaire du Cancer, Toulouse – Oncopole (IUCT-O), 1 avenue Irène Joliot-Curie, 31059 Toulouse Cedex 9, France

2   Biometric Unit, Institut Curie, Hospital Group, Paris, France

3   Data Center, Institut Curie, Hospital Group, Paris, France

4   Oncology Department, Institut Curie, Hospital Group, Paris, France

5   Université Paris Descartes, Paris, France

6   Radiotherapy Department, Institut Curie, Hospital Group, Paris, France

7   Oncology Department, Institut Claudius Regaud, Institut Universitaire du Cancer, Toulouse – Oncopole (IUCT-O), Toulouse, France

Background

The 5- or 10-year disease-free survival rate is commonly used as a criterion of success in non-metastatic cancer therapy. If this rate is estimated by Kaplan–Meier method [1], it is still required that a considerable proportion of all case have a sufficient follow-up. Pocock et al. [2] proposed to curtail Kaplan–Meier plot when only around 10–20 % are still in follow-up. Reducing the follow-up period time required to estimate long-term survival rates has become an important issue in oncology. The use of parametrical statistical models can be a solution but they have to be validated in practice. From the 50 s, several authors have worked on this topic. Considering only short-term survival data, the standard lognormal model was used to predict survival rates at 10, 15 years or more [3, 4]. The values predicted by the model were then compared to results observed at these time points. This strategy was also used to predict long-term survival of patients treated for cancer of the larynx [5] or inflammatory breast cancer [6].

The methodology used in these publications considers a single or a composite endpoint. However, in breast cancer, patients are at risk of multiple types of events. A patient can experience loco-regional recurrence, distant metastasis, contralateral cancer, second primary cancer or death. According to the nature of the event types, different treatment strategies may influence the time to occurrence of only certain event types. In practice, medical searchers may find interest in nature and time of a particular event, distant metastasis for example, and the other events are considered in competition with the event of interest. Kaplan–Meier estimator treated competing events as independent censorings and overestimates the proportion of distant metastasis. The cumulative incidence function (CIF) [7] takes into account the informative nature of the censoring and corresponds to the probability of occurrence of a particular event without the assumption of independence between event types. Recently, Jeong et al. proposed a parametric model which permits to directly model cumulative incidence. As this parametric method requires a minimum follow-up, it is therefore appropriate to propose an adaptation of this method by differentiating the types of event using the competing risks methodology. The direct approach proposed by Jeong and Fine was used [8]. The main objective of this paper was to evaluate the ability of this parametric model to predict the cumulative incidence rates at 10, 15, 20, 25 years using short-term data. Breast cancer data and statistical methodology are described in second section and results presented in third section will be discussed in fourth section.

Patients and methods

Study population

Data were obtained from the Institut Curie Breast cancer database. This database contains records from 32515 non-metastatic breast cancer patients treated between 1980 and 2005 at the Institut Curie. This study focused on patients without lymph node involvement (pN0) who underwent breast conservative surgery for a tumor size <5 cm (T1 or T2) and clinical size <3 cm. Patients not treated by radiotherapy and who had neoadjuvant or adjuvant chemotherapy were also excluded. This population includes 4761 patients treated between 1981 and 2000. Follow-up was until September 4, 2013.

Statistical analysis

Competing risks

A competing risk is defined as an event whose occurrence precludes or alters the probability of occurrence of a main

event under examination [9]. In this study, patients were considered at risk of three event types: loco-regional recurrence, distant metastasis and contralateral cancer or second cancer (another cancer in a different location), or death from any cause. The event of interest was the first event occurred.

It seems appropriate to distinguish loco-regional recurrence and distant metastasis, because these two event types are quite different in terms of predictive factors and survival prognosis of patients [10]. The third event is a composite event and includes contralateral cancer, second cancer and death from any cause. If a patient did not experience any event, it was censored at the date of last known contact.

In the setting of competing risks, the probability of failure overtime is described by the so-called cumulative incidence defined by

$$F_k(t) = P(T \le t, \Delta = k),$$

where T is the time between the diagnosis and the occurrence of the first event and $\Delta = 1, \ldots, K$ is the indicator of the event type.

Prediction models

Non-parametric

The Kalbfleisch and Prentice method [7] is widely used in the literature to summarize competing risks data. Let $t_1 < t_2 < \cdots < t_J$ represent J ordered times to first events. The non-parametric estimation of the cumulative incidence associated with an event type k ($k = 1, \ldots, 3$) is defined by

$$\hat{F}_k^{NP}(t) = \sum_{t \le t_j} \hat{S}(t_{j-1}) \times \hat{h}_k(t_j),$$

where $\hat{S}(t_{j-1})$ corresponds to event-free survival estimated by Kaplan–Meier method [1] at time t_{j-1} and $\hat{h}_k(t_j) = d_{kj}/n_j$ where d_{kj} is the number of type k events observed at time t_j and n_j is the number of patients at risk at time t_j.

Non-parametric confidence intervals are obtained using the log(-log) transformation proposed by Choudhury [11].

Parametric

Jeong and Fine [8] proposed a direct approach to model parametrically the cumulative incidence. Each cumulative incidence is directly modelized using a Gompertz distribution. The cumulative incidence of event k is defined by

$$F_k^P(t, \alpha_k, \beta_k) = 1 - \exp\left[\frac{\beta_k}{\alpha_k}(1 - \exp(\alpha_k t))\right],$$

where α_k is the shape parameter and β_k the scale parameter.

The Gompertz distribution is improper when $\alpha_k < 0$ and $\beta_k > 0$, leading to a plateau in the tail of the distribution which is particularly well suited to capture cure rates as it can be seen in breast cancer. The probability of experiencing the event of type k is $1 - \exp(\beta_k/\alpha_k)$. The parameters are estimated by the maximum-likelihood method. Here, patients were at risk of three event types, so a total of six parameters had to be estimated.

Using the model for prediction

The method used in order to evaluate the predictive ability of the model is illustrated by Fig. 1.

The study population was divided into four periods of 5 years where patients were treated, treatment intervals were defined as follows: 1981–1985, 1986–1990, 1991–1995, and 1996–2000. For each period, follow-up period was limited at $t_p = 1, 2, 3, 4,$ and 5 years, respectively. For example, a patient treated between 1981 and 1985 was followed at most until 1986, 1987, 1988, 1989, and 1990, respectively. After this period, it was censored if no events had occurred before. The parameters of the Jeong and Fine model were estimated using these short-term data (from $T = 0$ to $T = 5 + t_p$).

Non-parametric cumulative incidence functions estimated by Kalbfleisch and Prentice method using complete follow-up data (from $T = 0$ to $T = T_{max}$) were considered as the gold standard estimation. The long-term cumulative incidence functions predicted by Jeong and Fine model at 10, 15, 20, and 25 years were compared to this gold standard.

This procedure was iterated for each period of treatment and each follow-up period time.

The agreement between cumulative incidence rates predicted by the Jeong and Fine model and non-parametric estimation using Kalbfleisch and Prentice method at $T = 10, 15, 20,$ and 25 years was defined using validation criteria based on non-parametric confidence intervals proposed by Choudhury et al. [11] and defined by

$$\Delta : \hat{F}_k^{NP}(T)^{\exp\left[\frac{-c_{\alpha/2}\hat{\sigma}_k(T)}{\hat{F}_k^{NP}(T)\log(\hat{F}_k^{NP}(T))}\right]} \leq \hat{F}_k^{P}(T) \leq \hat{F}_k^{NP}(T)^{\exp\left[\frac{c_{\alpha/2}\hat{\sigma}_k(T)}{\hat{F}_k^{NP}(T)\log(\hat{F}_k^{NP}(T))}\right]},$$

where $c_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution and $\hat{\sigma}_k(T)$ is the square root of $\text{var}(\hat{F}_k^{NP}(T))$. Two validation criteria were computed, Δ_1 corresponding to 95 % non-parametric confidence interval ($\alpha = 1 - 0.95$) and Δ_2 corresponding to 99.9 % non-parametric confidence interval ($\alpha = 1 - 0.999$) .

$$\Delta_1 : CI_-^{95\%}\left(\hat{F}_k^{NP}(T)\right) \leq \hat{F}_k^{P}(T) \leq CI_+^{95\%}\left(\hat{F}_k^{NP}(T)\right)$$

$$\Delta_2 : CI_-^{99.9\%}\left(\hat{F}_k^{NP}(T)\right) \leq \hat{F}_k^{P}(T) \leq CI_+^{99.9\%}\left(\hat{F}_k^{NP}(T)\right)$$

The analysis was performed under R 2.15.1 software, four packages are also used: *survival, cmprsk, optimx,* and *numDeriv*.

Results

Patient's characteristics

A total of 4761 patients were extracted from the population database of breast cancer of Institut Curie. Patient characteristics are displayed in Table 1 for each period.

The median follow-up was 14.48 years in the overall population. The median age was 55 years (24–90). Majority of tumor size was lower than 2 cm (71.5 % T1 vs 28.5 % T2). About 1685 patients had tumor with grade I (37.3 %), 1776 grade II (39.3 %), and 1061 grade III (23.5 %). A majority of patients had hormone receptor-positive status (87.9 %). There were 673, 1178, 1382, and 1528 patients in each cohort, respectively. The median follow-up was 23.43 years for the patients treated between 1981 and 1985 and decreased until 11.18 years for the patients treated between 1996 and 2000. The median age was 49.5 years for the patients treated between 1981 and 1985 and increased until 58 years for the patients treated between 1996 and 2000. Proportion of T1 tumor and grade

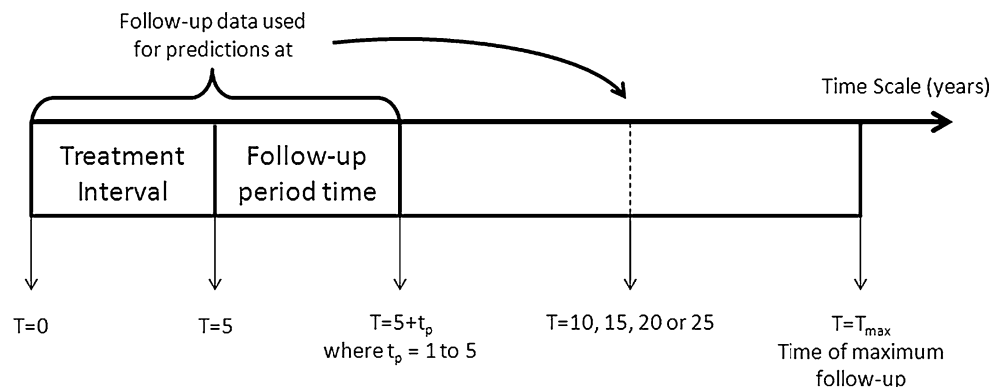**Fig. 1** Schematic diagram of evaluation predictive ability of the model

Table 1 Patient's characteristics

Treatment interval	1981–1985	1986–1990	1991–1995	1996–2000	Overall
No of patients	673	1178	1382	1528	4761
Follow-up (Y)					
Median	23.43	19.62	15.10	11.18	14.48
95 % CI	22.60–24.20	19.06–19.98	14.79–15.44	10.99–11.40	14.26–14.69
Age					
Median	49.5	53	56	58	55
Range	25–80	24–82	28–85	27–90	24–90
Tumor size					
T1	408 (60.6 %)	785 (66.6 %)	1031 (74.6 %)	1178 (77.1 %)	3402 (71.5 %)
T2	265 (39.4 %)	393 (33.4 %)	351 (25.4 %)	350 (22.9 %)	1359 (28.5 %)
Grade SBR					
I	144 (22.7 %)	375 (34.3 %)	493 (37.7 %)	673 (45.3 %)	1685 (37.3 %)
II	283 (44.6 %)	411 (37.6 %)	529 (40.4 %)	553 (37.2 %)	1776 (39.3 %)
III	207 (22.6 %)	308 (28.2 %)	286 (21.9 %)	260 (17.5 %)	1061 (23.5 %)
Missing	39	84	74	42	239
Hormonal receptors status					
ER− and PR−	97 (18.4 %)	146 (13.9 %)	141 (11.6 %)	98 (8.2 %)	482 (12.1 %)
ER+ and/or PR+	430 (81.6 %)	902 (86.1 %)	1077 (88.4 %)	1103 (91.8 %)	3512 (87.9 %)
Missing	146	130	164	327	767

Table 2 Number of each event types by treatment interval and follow-up period

Treatment interval	Event type	$t_p = 1$	$t_p = 2$	$t_p = 3$	$t_p = 4$	$t_p = 5$	T_{max}
1981–1985	Loco-regional recurrence	31	40	55	60	67	133
	Distant metastasis	38	44	51	57	60	99
	Contralateral/2nd cancer/death	27	32	40	48	50	171
1986–1990	Loco-regional recurrence	33	54	69	87	106	221
	Distant metastasis	54	75	90	101	111	144
	Contralateral/2nd cancer/death	41	65	75	89	105	276
1991–1995	Loco-regional recurrence	42	59	81	96	101	208
	Distant metastasis	43	58	69	79	84	118
	Contralateral/2nd cancer/death	55	65	79	94	122	273
1996–2000	Loco-regional recurrence	49	61	71	82	94	154
	Distant metastasis	37	47	55	61	68	86
	Contralateral/2nd cancer/death	58	75	99	123	146	233

I increased with time (60.6 % until 77.1 % and 22.7 % until 45.3 %). The same trend for proportion of hormone receptor-positive status (81.6 % until 91.8 %) was observed.

The number of each event types by treatment interval and follow-up period are displayed in Table 2.

In the first group, where patients were treated between 1981 and 1985, 133 loco-regional recurrences, 99 distant metastasis, and 171 contralateral cancer or second cancer or death were observed during the complete follow-up (from 1981 to today). If the follow-up was limited at 1 year (from 1981 to 1986), only 31 loco-regional recurrences, 38 distant metastasis, and 27 contralateral cancer or second cancer or death were observed. If $t_p = 5$ years, 67 loco-regional recurrences, 60 distant metastasis, and 50 contralateral cancer or second cancer or death were observed from 1981 to 1990.

Prediction results

Table 3 displays the cumulative incidence rates predicted by the model and non-parametric estimates with 95 and 99.9 % confidence intervals at 10, 15, 20, and 25 years for each event types using data from the first treatment interval period

Table 3 Cumulative incidence rates of loco-regional recurrence, distant metastasis, and contralateral/2nd cancer/death for the 1981–1985 treatment period

	$T = 10$	$T = 15$	$T = 20$	$T = 25$
Loco-regional recurrence				
Non-parametric estimate	0.126	0.149	0.182	0.226
95 % CI	0.102–0.153	0.123–0.178	0.152–0.213	0.190–0.264
99.9 % CI	0.088–0.172	0.107–0.198	0.134–0.235	0.167–0.290
Jeong and Fine				
$t_p = 1$	0.124	0.168	0.205	0.235
$t_p = 2$	0.127	0.176	0.216	0.250
$t_p = 3$	0.146	0.203	0.251	0.293
$t_p = 4$	0.134	0.184	0.225	0.259
$t_p = 5$	0.135	0.189	0.237	0.279
Distant metastasis				
Non-parametric estimate	0.110	0.138	0.147	0.154
95 % CI	0.087–0.135	0.113–0.166	0.121–0.176	0.126–0.183
99.9 % CI	0.074–0.153	0.098–0.186	0.105–0.196	0.109–0.205
Jeong and Fine				
$t_p = 1$	0.146	0.193	0.228	0.254
$t_p = 2$	0.131	0.169	0.196	0.216
$t_p = 3$	0.118	0.144	0.159	0.168
$t_p = 4$	0.122	0.156	0.179	0.196
$t_p = 5$	0.109	0.130	0.141	0.147
Contralateral/2nd cancer/death				
Non-parametric estimate	0.113	0.176	0.240	0.292
95 % CI	0.090–0.139	0.148–0.207	0.206–0.275	0.253–0.332
99.9 % CI	0.077–0.158	0.130–0.226	0.185–0.299	0.227–0.360
Jeong and Fine				
$t_p = 1$	0.105	0.139	0.165	0.185
$t_p = 2$	0.099	0.132	0.158	0.178
$t_p = 3$	0.099	0.128	0.148	0.163
$t_p = 4$	0.107	0.144	0.173	0.197
$t_p = 5$	0.095	0.119	0.136	0.147

(1981–1985). Graphical representation of these results is shown in Figs. 2a–c. Results obtained from the three other groups of treatment periods (1986–1990, 1991–1995, and 1996–2000) are available in the supplementary materials. The values of CIF in the body of Table 3 are shaded when they satisfy the validation criteria Δ_1 or Δ_2.

Loco-regional recurrence

Non-parametric estimates of cumulative incidence functions of loco-regional recurrence were equal to 0.126, 0.149, 0.182, and 0.226 at 10, 15, 20, and 25 years, respectively. Cumulative incidence rates at 10 years predicted by Jeong and Fine model satisfied Δ_1 (95 % CI = [0.102; 0.153]) whatever the follow-up period time. They were equal to 0.124, 0.127, 0.146, 0.134, and 0.135 with $t_p = 1$, 2, 3, 4, and 5 years, respectively. Predictions at

15 years satisfied Δ_1 with limited follow-up at 1 or 2 years (0.168 and 0.176) and were slightly overestimated with $t_p = 3$, 4, or 5 years (0.203, 0.184, and 0.189). At 20 and 25 years, predictions were also slightly overestimated except that with limited follow-up at 1 year (0.205 and 0.235) and $t_p = 2$ or 4 years for 25 years prediction (0.250 and 0.259). A maximum difference of 5.5 % was observed between non-parametric estimation and prediction rates at 20 or 25 years with follow-up limited at 4 or 5 years (0.237 − 0.182 = 0.055).

All prediction rates satisfied Δ_2 except prediction rates at 15, 20, and 25 years with $t_p = 3$ years and prediction rate at 20 years with $t_p = 5$ years.

For the other treatment periods, prediction rates were generally slightly underestimated with follow-up limited at 1, 2, or 3 years. But all prediction rates satisfied Δ_1 or Δ_2 with follow-up limited at 3, 4, or 5 years.
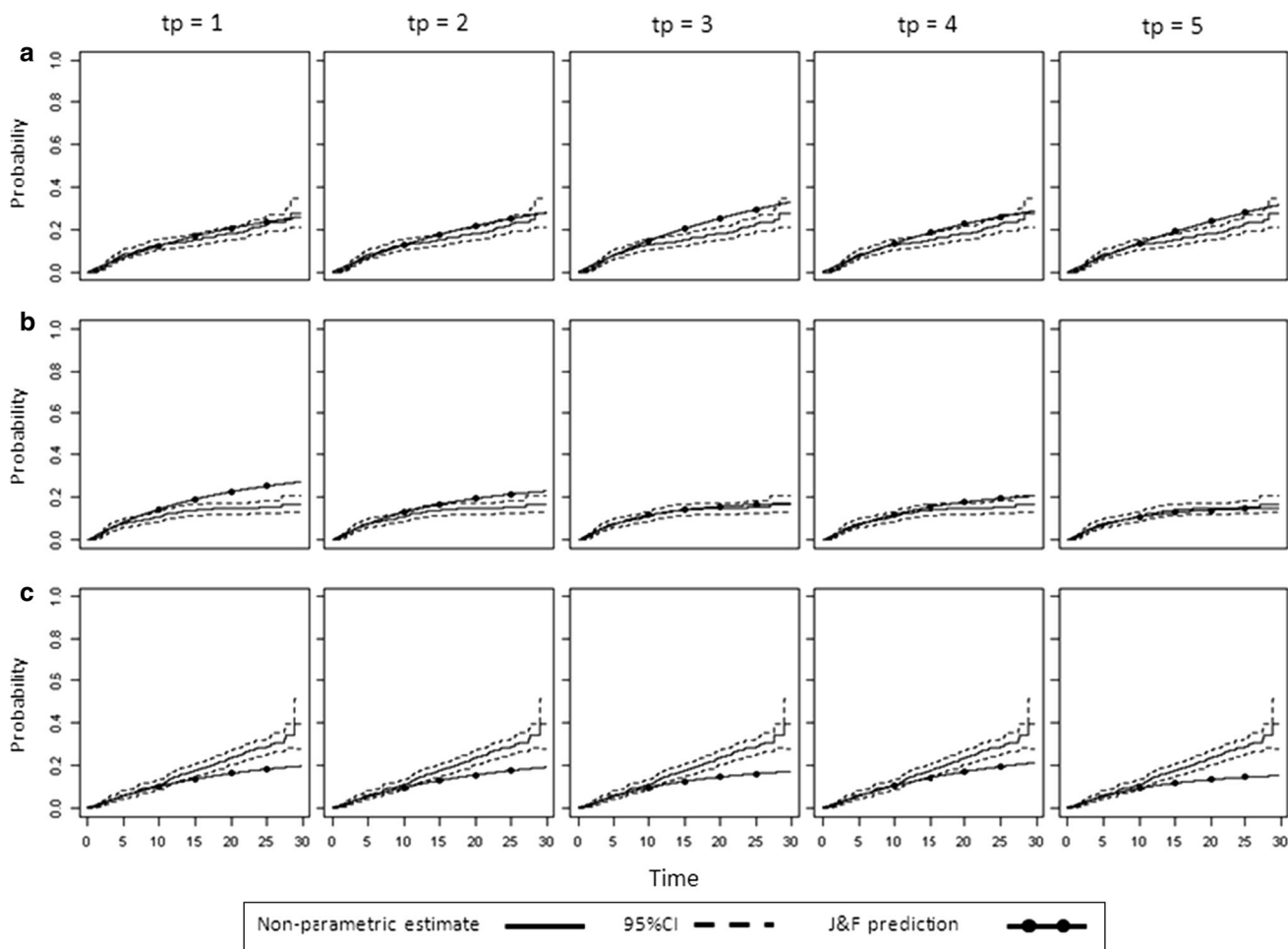
Fig. 2 Cumulative incidence rates of loco-regional recurrence (**a**), distant metastasis (**b**), and contralateral/2nd cancer/death (**c**) for the 1981–1985 treatment period

Distant metastasis

Non-parametric estimates of cumulative incidence functions of distant metastasis were equal to 0.110, 0.138, 0.147, and 0.154 at 10, 15, 20, and 25 years, respectively. Cumulative incidence rates at 10 years predicted by Jeong and Fine model satisfied Δ_1 (95 % CI = [0.087; 0.135]) whatever the follow-up period time except with follow-up limited at 1 year (0.146). Predictions at 15 years satisfied Δ_1 with limited follow-up at 3, 4, or 5 years (0.144, 0.156, and 0.130) and were slightly overestimated with $t_p = 1$ or 2 years (0.193 and 0.169). At 20 and 25 years, predictions were also slightly overestimated except that with $t_p = 3$ or 5 years for 20 years (0.159 and 0.141) and 25 years predictions (0.168 and 0.147). A maximum difference of 4.2 % was observed between non-parametric estimation and prediction rates at 20 or 25 years with follow-up limited at 4 or 5 years (0.196 − 0.154 = 0.042).

All prediction rates satisfied Δ_2 except prediction rates at 15, 20, 25 years with $t_p = 1$ years and prediction rate at 25 years with $t_p = 2$ years.

For the other treatment periods, prediction rates were also generally overestimated with follow-up limited at 1 or 2 years. But all prediction rates satisfied Δ_1 or Δ_2 with follow-up limited at 3, 4, or 5 years. Except for the 1986–1990 period where prediction rates were overestimated even if follow-up was limited at 5 years.

Contralateral or second cancer or death

Non-parametric estimates of cumulative incidence functions of contralateral or second cancer or death were equal to 0.113, 0.176, 0.240, and 0.292 at 10, 15, 20, and 25 years, respectively. Cumulative incidence rates at 10 years predicted by Jeong and Fine model satisfied Δ_1 (95 % CI = [0.090; 0.139]) whatever follow-up period time. They were equal to 0.105, 0.099, 0.099, 0.107, and 0.095 with $t_p = 1, 2,$

3, 4, and 5 years, respectively. However, predictions at 15, 20, and 25 years were all underestimated with any follow-up period time. They were equal or lower than 0.144, 0.173, and 0.197 at 15, 20, and 25 years, while lower bounds of 95 % non-parametric confidence intervals were equal to 0.148, 0.206, and 0.253, respectively.

Only prediction rates at 15 years with follow-up limited at 1, 2 or 4 years which were equal to 0.139, 0.132 and 0.144 satisfied Δ_2 (99.9 % CI = [0.130; 0.226]).

For the other treatment periods, prediction rates were generally also underestimated whatever follow-up period time.

Similar results for the three event types were obtained in the subpopulation with hormone receptor-positive status (data not shown). Due to the small number of patients with hormone receptor-negative status, it was not possible to perform analysis in this subpopulation.

Discussion

Competing risks methodology has been developed for more than 30 years and this setting is increasingly used in cancer research. The aim of this paper was to evaluate the ability of competing risks parametric model to predict long-term cumulative incidences using short-term breast cancer data. A good prognosis population was selected for predictions in order to have a homogeneous sample in terms of risk of disease recurrence and overall survival. The direct approach proposed by Jeong and Fine seems to provide a useful alternative to non-parametric method of cumulative incidence estimation even when follow-up data are sufficiently extensive. Predictive ability of this model on breast cancer data was generally good considering the short follow-up period time used for estimation.

For the loco-regional recurrence risk, predictions were generally accurate even if follow-up period was very short (1 or 2 years). With follow-up limited at 3 years, Jeong and Fine model overestimated cumulative incidence rates but predictions were better with a longer follow-up period (i.e., 4 or 5 years). For the distant metastasis risk, predictions generally satisfied the validation criteria, except with follow-up limited at 1 or 2 years where long-term predictions were overestimated. Saphner et al. [12] highlighted the presence of a peak hazard of recurrence after several years of follow-up and decreasing thereafter in breast cancer after primary therapy. This phenomenon can explain overestimation of prediction rates of loco-regional recurrence and distant metastasis. Indeed, if follow-up period is not sufficiently large to capture the peak hazard of recurrence, cumulative incidence rates can be overestimated. It therefore would be preferable to fit the model using sufficiently large follow-up period in order to averaging this

peak hazard of recurrence. In the second group where patients were treated between 1986 and 1990, predictions rates of distant metastasis were highly overestimated even if follow-up was limited at 5 years. Follow-up were extended to 6 or 7 years in order to capture the peak hazard of the first years, and prediction rates satisfied validation criteria (see Supplementary materials). Sensitivity analyses were also performed on treatment period (2 and 3 years) to evaluate the performance of the technique with fewer events. Due to the smaller number of patients and a shorter follow-up period, long-term predictions were less accurate (data not shown).

95 % non-parametric confidence interval (Δ_1) was considered as validation criteria for the predictive ability of the model. Mould [5] suggested relaxing the validation criteria for long-term prediction at 20 and 25 years. Cumulative incidence functions were overall well predicted by the model at 10 and 15 years but they were less accurate in the long term according to Δ_1. Indeed, predictions were calculated using at most 10 years of follow-up data making at least 10 and 15 years of follow-up saved rather than non-parametric method for predictions at 20 and 25 years, respectively. It seems reasonable, given the number of years saved, to apply this principle and use a wider non-parametric confidence interval as validation criteria for long-term predictions. 99.9 % non-parametric confidence interval (Δ_2) was also used. In the first treatment group, width of Δ_2 at 20 years (resp. 25 years) was 10.1 % (resp. 12.3 %) for the loco-regional recurrence and 9.1 % (resp. 9.6 %) for the distant metastasis. Most of the prediction rates with follow-up limited at 4 or 5 years satisfied Δ_2, except prediction rate at 20 years with $t_p = 4$ for the loco-regional recurrence which was equal to 0.237 and 99.9 % CI = [0.134; 0.235]. Prediction rates at 20 or 25 years had a maximum difference of around 6 % compared to the gold standard estimation. The same trend for the other treatment periods was observed (see Supplementary materials). It is also possible to apply these validation criteria at any time provided that there were enough patients still at risk at this time in order to have accurate non-parametric estimates. As pointwise confidence interval is valid for a single fixed time, an alternative approach is the use of confidence bands. Upper and lower confidence band guarantee, with a confidence level, the cumulative incidence function that falls within the band. This could allow the possibility to evaluate the predictive ability of the model over time and not at a specified time points.

For the third event type (contralateral cancer or second cancer or death), predictions of cumulative incidence were less accurate than the two first event types. The heterogeneity of this composite event including contralateral cancer, second cancer, and death may explain this lack of precision. Furthermore, this model is built to be applied on

a distribution where hazard decreases over time. In this study, death from any cause was considered and not only death from cancer. It is quite unusual to die from cancer without experiencing a local or distant relapsed first, so it is reasonable to think that most deaths observed as first event were not related to cancer. Given the age at the end of treatment of patients (49.5 years for the 1981–1985 period), the instantaneous hazard of death increased with time instead of decreasing as assumed by the model. This feature could also explain the underestimations of cumulative incidence of contralateral or second cancer or death at long term. If only death related to cancer would be considered instead of death from any cause, predictions of cumulative incidence of the third event type would be more accurate. Several authors [8, 13] recommend testing the goodness of fit of the data to the model before using it. The modified Gompertz model is only applicable when the hazard function decreases with time. This is why this model is more suitable to predict cumulative incidence of a particular event like loco-regional recurrence or distant metastasis than the third event where death not related to cancer was considered. In fact at infinity, all patients cured from the disease will die from other cause.

This is one of the limits of this direct estimation approach, the fitted event-free survivor function is not constrained to be a proper survivor function, i.e., the sum of the cumulative incidence functions at infinity equals 1. More recently, Shi et al. [14] proposed a parameterization that explicitly accounts this constraint [15]. This merits further investigation.

Another important aspect in oncology is the probability of cure which commonly names as the cure rate. In fact, recent advances in cancer treatment have modified the course of some cancer types and some patients will never relapse after treatment [16]. The term of cure can be used only when data present a large number of long-term survivor. Different models address the problem of cure rate estimation in the analysis of survival data [3, 17, 18] but few attentions were given in the literature concerning the evaluation of cure rate in the presence of competing risks data [19, 20]. Future work need to evaluate statistical tools to analyze competing risk data in the presence of cure fraction.

The Jeong and fine prediction model provides a useful addition to the Kalbfleisch and Prentice non-parametric method for the estimation of cumulative incidence. Using this model, it is possible to save a significant number of follow-up years for the estimation of long-term data. This could result in more rapid advances in breast cancer treatment and have the potential benefit of a reduction of cost. However, this prediction model can only be used in clinical practice when it is fully validated for a cancer site.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. J Am Stat Assoc 53(282):457–481
2. Pocock SJ, Clayton TC, Altman DG (2002) Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. Lancet 359(9318):1686–1689
3. Boag JW (1949) Maximum likelihood estimates of the proportion of patients cured by cancer therapy. J R Stat Soc Ser B Methodol 11(1):15–53
4. Mould RF, Boag JW (1975) A test of several parametric statistical models for estimating success rate in the treatment of carcinoma cervix uteri. Br J Cancer 32(5):529–550
5. Mould RF, Lederman M, Tai P, Wong JKM (2002) Methodology to predict long-term cancer survival from short-term data using Tobacco Cancer Risk and Absolute Cancer Cure models. Phys Med Biol 47(22):3893
6. Tai P, Yu E, Shiels R, Pacella J, Jones K, Sadikov E, Mahmood S (2005) Short-and long-term cause-specific survival of patients with inflammatory breast cancer. BMC Cancer 5(1):137
7. Kalbfleisch JD, Prentice RL (2011) The statistical analysis of failure time data, vol 360. Wiley, Hoboken
8. Jeong J-H, Fine J (2006) Direct parametric inference for the cumulative incidence function. J R Stat Soc Ser C Appl Stat 55(2):187–200
9. Gooley TA, Leisenring W, Crowley J, Storer BE et al (1999) Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. Stat Med 18(6):695–706
10. Komoike Y, Akiyama F, Iino Y, Ikeda T, Akashi-Tanaka S, Ohsumi S, Kusama M, Sano M, Shin E, Suemasu K, Sonoo H, Taguchi T, Nishi T, Nishimura R, Haga S, Mise K, Kinoshita T, Murakami S, Yoshimoto M, Tsukuma H, Inaji H (2006) Ipsilateral breast tumor recurrence (IBTR) after breast-conserving treatment for early breast cancer. Cancer 106(1):35–41
11. Choudhury JB (2002) Non-parametric confidence interval estimation for competing risks analysis: application to contraceptive data. Stat Med 21(8):1129–1144
12. Saphner T, Tormey DC, Gray R (1996) Annual hazard rates of recurrence for breast cancer after primary therapy. J Clin Oncol 14(10):2738–2746
13. Somda SM, Leconte E, Boher J-M, Asselain B, Kramar A, Filleron T (2014) Optimal scheduling of post-therapeutic follow-up of patients treated for cancer for early detection of relapses. Stat Methods Med Res 0962280214524178. http://www.ncbi.nlm.nih.gov/pubmed/24567440
14. Shi H, Cheng Y, Jeong J-H (2013) Constrained parametric model for simultaneous inference of two cumulative incidence functions. Biom J 55(1):82–96
15. Moreno-Betancur M, Rey G, Latouche A (2015) Direct likelihood inference and sensitivity analysis for competing risks regression with missing causes of failure. Biometrics 71(2):498–507

16. Slamon D, Eiermann W, Robert N, Pienkowski T, Martin M, Press M, Mackey J, Glaspy J, Chan A, Pawlicki M et al (2011) Adjuvant trastuzumab in HER2-positive breast cancer. N Engl J Med 365(14):1273–1283

17. Sy JP, Taylor JM (2000) Estimation in a Cox proportional hazards cure model. Biometrics 56(1):227–236

18. Andersson TM, Dickman PW, Eloranta S, Lambert PC (2011) Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. BMC Med Res Methodol 11(1):96

19. Choi KC, Zhou X (2002) Large sample properties of mixture models with covariates for competing risks. J Multivar Anal 82(2):331–366

20. Nicolaie MA, van Houwelingen HC, Putter H (2010) Vertical modeling: a pattern mixture approach for competing risks modeling. Stat Med 29(11):1190–1205