European Actuarial Journal
https://doi.org/10.1007/s13385-024-00403-6

ORIGINAL RESEARCH PAPER

Right to be forgotten for mortgage insurance issued to cancer survivors: critical assessment and new proposal

Antoine Soetewey[1] · **Catherine Legrand**[1] · **Michel Denuit**[1] · **Geert Silversmit**[2]

Received: 4 December 2023 / Revised: 24 September 2024 / Accepted: 3 October 2024

Abstract
Soetewey et al. (Eur Actuar J 11(1):135–160, 2021) proposed to determine the waiting period opening the right to be forgotten (RTBF) as the time after diagnosis needed for the premium to revert back to some acceptable level expressed by means of regulatory life tables. However, this approach requires data up to 30 years after diagnosis (10 years of standard RTBF plus the typical duration of the loan), or extrapolating the results up to that time horizon. When survival statistics are only available over a shorter duration, it turns out that the results may strongly depend on the extrapolation method. This is why an alternative method is proposed here, based on a constraint imposed to the premium. This constraint is then transposed into a target on the conditional observed survival and the waiting period follows. For the sake of robustness, results obtained with the proposed approach are compared to results obtained with Kaplan–Meier estimate taken as a non-parametric reference. Furthermore, the paper investigates the impact of the stage of the tumor at diagnosis on waiting periods.

Keywords Term insurance · Impaired lives · Waiting period · Home loan · Cancer stage at diagnosis

1 Introduction and motivation

Outstanding balance insurance is generally required by lenders to secure their loans. The borrower is the insured life: if he or she dies before the loan has been fully repaid then the insurer pays a death benefit corresponding to the balance of the loan. As with any other term life insurance product, applicants with poor health conditions may be denied insurance or charged increased amounts of premium compared to standard

✉ Antoine Soetewey
antoine.soetewey@uclouvain.be

[1] Institute of Statistics, Biostatistics and Actuarial Sciences, Louvain Institute of Data Analysis and Modeling in Economics and Statistics UCLouvain, Louvain-la-Neuve, Belgium

[2] Belgian Cancer Registry, Brussels, Belgium

Published online: 04 November 2024

conditions. In extreme cases, this may prevent them from accessing property or develop their business project. For this reason, several EU countries passed laws to ease access to mortgage insurance for long-term disease survivors.

The "right to be forgotten" (henceforth abbreviated as RTBF) emerged under strict European data protection regulations, which allow individuals to request the deletion of their personal data collected online or through social networks. This RTBF is distinct from its application in insurance contexts, where it pertains to the handling of health information for insurance applicants. In insurance, the RTBF allows individuals with a history of certain health conditions, such as cancer, to not disclose their previous illness after a specified period (or to declare it but forbidding the insurer to take this information into account, as specified in the Belgian law), thus improving their access to insurance coverage.

The first initiative in this direction dates back to 2007, when France launched the AERAS Convention (AERAS is the acronym for "*s'Assurer et Emprunter avec un Risque Aggravé de Santé*" in French, which could be translated as "insuring and borrowing with an aggravated health risk"). This agreement, signed by the public authorities, banking and insurance sectors, and patients' and consumers' associations purposed to allow people cured from cancer or suffering certain chronic diseases to access insurance comprising benefits in case of death or disability, as well as to guaranteed income insurance. Considering long-term cancer survivors, the AERAS Convention included a RTBF, that is, the right for an insurance applicant not to declare a previous cancer after a period of 10 years starting at the end of the therapeutic protocol (reduced to 5 years for pediatric cancers). These periods of 10 and 5 years start from the date of the end of the therapeutic treatment, in absence of relapse within this period.

In Belgium, it was only in 2019 that the RTBF entered the insurance law. Based to a large extent on the reference tables published in the AERAS Convention, a Royal Decree dated May 26, 2019 lists certain types of cancer for which, depending upon entry criteria (such as cancer stage or age), the standard waiting period of 10 years from the end of active treatment is reduced. The RTBF has recently been adapted in Belgium, again following similar changes in France. As from November 2022, the standard waiting period opening the RTBF has been shortened from 10 to 8 years, and it has been adopted that it will be reduced to 5 years as from January 2025. Moreover, also as from November 2022, the period is shortened to 5 years for cancer survivors who have been cured before the age of 21.

The RTBF has now also been installed through an agreement in Luxembourg, and through a legal framework in The Netherlands, Portugal, Romania, and more recently in Italy, Spain and Cyprus. It is being debated and advocated for at the European level to expand to the other EU countries as well. There are some ongoing discussions between Insurance Europe, the European Commission and the European Parliament on a possible EU-wide RTBF for cancer survivors. See for instance Scocca and Meunier [18, 19]. Moreover, the European Commission's "Europe's Beating Cancer Plan" includes actions to enhance cancer survivors' access to financial services, including insurance. This plan proposes the development of a self-regulated Code of Conduct and considers long-term solutions for improving access to financial services for cancer survivors.

This paper aims to contribute to this evolution by proposing an actuarially sound methodology to evaluate a technically correct waiting period opening the RTBF. It starts with a critical assessment of the approaches proposed by Soetewey et al. [20] and by Van Ginckel et al. [24]. It turns out that the results obtained from the method proposed by Soetewey et al. [20] strongly depend on the extrapolation method. This is precisely shown in this paper, by modifying the extrapolation method and ending up with different waiting periods for some cancer types. This is clearly not acceptable in the context of the RTBF. To be precise, the problem is not with the method proposed by Soetewey et al. [20] but comes from the limited follow-up period for patients in some cancer registries, including the Belgian one. This requires extrapolation to longer times since diagnosis and this step may induce higher uncertainty. If the length of the follow-up is enough, the method proposed by Soetewey et al. [20] remains actuarially sound.

We also consider the approach proposed by Van Ginckel et al. [24], which applies a biostatistical approach based on an arbitrary cut-off of 0.99 for the conditional net survival to propose reduced waiting periods for breast cancer. In Sect. 4.2, we show that, despite the apparent closeness to general population mortality in terms of survival, their method results in one-year death probabilities up to 10 times higher compared to general population at young adult ages. It is clear that such excess mortality cannot be absorbed by mortgage insurance market without increasing premiums at standard conditions.

There is thus a need for another approach, escaping the problem faced with extrapolation in case of limited follow-up and controlling the resulting premium compared to some market reference. In this paper, we impose a constraint to the premium and transpose it into a target on the conditional observed survival probabilities. The main assumption retained throughout this paper is that mortality for cancer patients temporarily peaks after diagnosis before reverting back to a level comparable to the general population for survivors. We will demonstrate in this paper how the length of the waiting period opening the RTBF can be derived from the comparison of the conditional one-year survival probabilities of cancer patients with the corresponding probabilities at general population level. The main advantage is that the time from which the RTBF can be exercised can be estimated from the available data only, without the need to extrapolate mortality rates beyond 10 years. While cancer stage at diagnosis has not been taken into account in Soetewey et al. [20], the present paper studies how the length of the waiting period opening the RTBF varies according to the extent of the tumor at diagnosis.

The remainder of this paper is structured as follows. Section 2 describes the mortgage insurance product considered in this paper. Section 3 presents the data used to perform the present study. Section 4 critically assesses the methods proposed by Soetewey et al. [20] and by Van Ginckel et al. [24]. It is shown that the chosen extrapolation method for limited follow-up impacts on the length of the resulting waiting period opening the RTBF. Our alternative approach is detailed in Sect. 5, and results obtained with our approach are compared with results obtained via a method based on the non-parametric Kaplan–Meier estimator to demonstrate that the proposed approach is trustworthy. Waiting periods are then derived from the comparison between cancer patients' conditional one-year survival probabilities and conditional one-year survival

probabilities of the general population. In Sect. 6, we illustrate these analyses considering Belgian data on melanoma, thyroid and female breast cancers according to the stage of tumor at diagnosis. The final Sect. 7 concludes the paper with a discussion.

2 Mortage insurance

The proposed approach is based on a representative mortgage insurance contract for the market under consideration. In this paper, we work with a simplified example which could be an appropriate starting point. Precisely, we consider a mortgage insurance applicant aged x borrowing an amount of capital κ at annual interest rate r for a duration n. The capital is reimbursed by constant yearly installments over the n years. The borrower pays the amount

$$\frac{\kappa}{a_{\overline{n}|}}, \text{ where } a_{\overline{n}|} = \sum_{k=1}^{n} \frac{1}{(1+r)^k},$$

back to the lender. Working with annual repayments compared to monthly ones is conservative from the insurer's point of view.

At time s, the amount of the loan that has not yet been amortized is denoted as c_s. Let $\lfloor s \rfloor$ denotes the integer part of $s \in [0, n]$, that is, the largest integer smaller than, or equal to s. At time s, $0 < s \leq n$, the present value of future payments is

$$c_s = c_{\lfloor s \rfloor}(1+r)^{s-\lfloor s \rfloor}$$

where $c_{\lfloor s \rfloor}$ is the outstanding balance of the loan at time $\lfloor s \rfloor$, right after the yearly installment has been paid, given by

$$c_{\lfloor s \rfloor} = \kappa \frac{a_{\overline{n-\lfloor s \rfloor}|}}{a_{\overline{n}|}}.$$

The loan is secured by a mortgage insurance, repaying the lender the amount c_s in case the policyholder dies at time s, $0 < s \leq n$. The net single premium is the expected present value (henceforth abbreviated as EPV) of insurance benefits, that is,

$$\pi_0 = \int_0^n {}_s p_x \mu_{x+s} c_s (1+i)^{-s} \mathrm{d}s$$

where i is the technical interest rate for the insurance contract, ${}_s p_x$ is the s-year survival probability for a policyholder aged x at policy issue and μ_{x+s} is the hazard, or force of mortality, at attained age $x+s$. Note that hazard and force of mortality are the same. For consistency, from now on, hazard will always be used in this paper. In accordance with actuarial notation, we denote as p_y the one-year survival probability at integer age y (that is, the probability of being alive at age $y+1$ given that the individual is alive at age y). Note that y is introduced to differentiate from x which refers to the

age at policy issue. In this sense, although both are equal in terms of value, y refers to a generic age while x refer to the specific age at policy issue.

Premium calculation is often based on regulatory or experience life tables. In this paper, we consider that standard conditions correspond to premiums computed according to the Belgian regulatory life table XK applying to insurance products comprising benefits in case of death (formally, XK defines minimum premium amount for policies with a positive sum at risk). This life table is widely adopted by Belgian insurers. It is known to be conservative and to generate a relatively high safety loading. Insurers are also allowed to use experience life tables available from the website of the National Bank of Belgium (NBB). These life tables reflect the mortality observed on the market, within portfolios of companies controlled by NBB. There is no safety loading and insurers are only allowed to apply premium rates resulting from NBB tables for relatively short periods of time (5 years, and then rates are subject to revision in case the observed mortality on the market changes over time). In this paper, we only consider the XK life table for premium calculation since these tables can be guaranteed for the whole contract duration and their conservatism better reflects increased mortality levels due to the disease.

The XK life table published in a Royal Decree does not distinguish between male and female policyholders, in accordance with EU anti-discrimination directive. For this reason, the entire analysis is conducted in this paper by pooling male and female mortality data. Also, it only gives one-year survival probabilities p_y at integer ages y. In this paper, we work under piecewise constant hazard, assuming that

$$\mu_{y+s} = \mu_y = -\ln p_y \text{ for all } 0 \leq s < 1 \text{ and integer } y.$$

Let us compute π_0 under this assumption. To this end, we split the integral to get

$$\begin{aligned}
\pi_0 &= \sum_{k=0}^{n-1} \int_k^{k+1} {}_s p_x \mu_{x+s} c_s (1+i)^{-s} \mathrm{d}s \\
&= \sum_{k=0}^{n-1} {}_k p_x \int_0^1 {}_s p_{x+k} \mu_{x+k+s} c_k (1+r)^s (1+i)^{-k-s} \mathrm{d}s \\
&= \sum_{k=0}^{n-1} {}_k p_x (1+i)^{-k} c_k \mu_{x+k} \int_0^1 {}_s p_{x+k} (1+r)^s (1+i)^{-s} \mathrm{d}s.
\end{aligned}$$

Now,

$$\begin{aligned}
\int_0^1 {}_s p_{x+k} (1+r)^s (1+i)^{-s} \mathrm{d}s &= \int_0^1 \exp\Big(-s\big(\mu_{x+k} - \ln(1+r) + \ln(1+i)\big)\Big) \mathrm{d}s \\
&= \frac{1 - \exp\big(-\mu_{x+k}\big)\frac{1+r}{1+i}}{\mu_{x+k} - \ln(1+r) + \ln(1+i)},
\end{aligned}$$

so that we finally get

$$\pi_0 = \sum_{k=0}^{n-1} {}_k p_x (1+i)^{-k} c_k \mu_{x+k} \frac{1 - \exp\left(-\mu_{x+k}\right)\frac{1+r}{1+i}}{\mu_{x+k} - \ln(1+r) + \ln(1+i)} \tag{2.1}$$

where ${}_0 p_x = 1$ and for $k \geq 1$,

$${}_k p_x = \prod_{j=0}^{k-1} p_{x+j} = \exp\left(-\sum_{j=0}^{k-1} \mu_{x+j}\right).$$

3 Data

The data available from the Belgian Cancer Registry (BCR) are considered in this paper. The BCR is a national population-based cancer registry collecting data on all new cancer cases diagnosed in Belgium since the incidence year 2004. Cancer registration has been made compulsory by law since 2006 in Belgium. The vital status is derived from linkage with the Belgian Crossroads Bank for Social Security up to April 11, 2022 and quality controls are performed regularly by BCR, ensuring the continuity and completeness of cancer registration in the country. More information can be found on the BCR website, at http://www.kankerregister.org.

To illustrate our work, three cancer types are considered: melanoma (ICD-10 C43), thyroid (ICD-10 C73) and female breast (ICD-10 C50) cancer (only female breast cancer is considered as there are too few registrations for male breast cancer). These three cancer sites have been selected to evaluate the proposed method in different scenarios. Melanoma and thyroid cancer patients are known to have a limited excess hazard compared to the general population [20]. The situation for female breast cancer patients is different with usually a high yearly survival probability in the first years after the date of diagnosis before it eventually decreases due to late cancer recurrences. Moreover, it is known that mortality for patients diagnosed with any of these three cancer types varies with time since diagnosis [21], yielding appropriate illustrations of the right to be forgotten.

For these applications, our analyses are also limited to patients aged 20 to 69 at time of diagnosis for two main reasons. First, childhood cancers can be seen as a category of cancer on their own, and are often studied separately because they greatly differ from adult cancers. Second, the RTBF mainly concerns young adults and active life.

Out of a total of 161,007 tumors, melanoma, thyroid and breast cancer represent, respectively, 29,213 (18.1%), 12,241 (7.6%) and 119,553 (74.3%) cases diagnosed between 2004 and 2020. Patients were followed-up until April 11, 2022, resulting in a follow-up ranging from 2 to 18 years. Only one record per patient (with the earliest incidence date) within each cancer site was kept for patients with multiple primary diagnoses. A minority of patients without national security number were excluded from the analysis. Patients lost to follow-up (mostly due to moving abroad) and patients

Table 1 Number of persons diagnosed with melanoma, thyroid and female breast cancer in Belgium between 2004 and 2020 (BCR data) by sex, site and age group, together with the percentage of lost to follow-up and the number of deaths

Sex	Cancer site	Age at diagnosis	Lost to follow-up (%)	Number of included cases	Number of deaths
Men	Melanoma	20–34	3.72	969	94
		35–49	2.66	3,266	404
		50–69	1.70	7460	1583
Total				11,695	2,081
Men	Thyroid	20–34	4.10	366	6
		35–49	3.12	961	67
		50–69	2.14	1773	379
Total				3,100	452
Women	Melanoma	20–34	3.62	2488	78
		35–49	1.47	6137	382
		50–69	1.35	8893	1112
Total				17,518	1572
Women	Thyroid	20–34	3.80	1607	14
		35–49	2.67	3449	107
		50–69	2.06	4085	484
Total				9,141	605
Women	Breast	20–34	2.76	3112	502
		35–49	1.78	32,743	4058
		50–69	1.31	83,698	15,946
Total				119,553	20,506

still alive at the end of the follow-up period were treated as censored observations. Censoring is assumed to be uninformative.

Table 1 summarizes the number of included cases, number of deaths and percentage of lost to follow-up before April 11, 2022 per type of cancer, sex and age group. The fraction of patients lost to follow-up per subgroup varied from 1.31% for women with breast cancer aged 50–69 to 4.1% for male thyroid cancer patients aged 20–34. The total fraction of patients lost to follow-up cases, regardless of sex, site or age group was 1.64%. Moreover, mean age at diagnosis was 50.5 years (standard deviation 12.1), 48.1 years (standard deviation 12.4) and 54.6 years (standard deviation 9.5) for melanoma, thyroid and breast cancer, respectively.

For the mortality in the general population, Belgian population life tables are available from Statbel (the Belgian statistical office) and can be freely downloaded from the website http://www.statbel.fgov.be.

4 Critical assessment

In this section, we revisit previous studies by Soetewey et al. [20] and Van Ginckel et al. [24] to underline their possible shortcomings.

4.1 Impact of extrapolation in case of limited follow-up

In this paper, we analyze survival time from diagnosis for cancer patients according to a number of covariates summarized into the vector z. Specifically, T denotes the remaining lifetime at diagnosis, so time from diagnosis to death. Given z, T has probability density function $f(\cdot|z)$, distribution function $F(\cdot|z)$, survival function $S(\cdot|z) = 1 - F(\cdot|z)$, and hazard $\lambda(\cdot|z) = f(\cdot|z)/S(\cdot|z)$. Contrarily to insurance studies, T denotes the remaining lifetime after diagnosis and age at diagnosis is included as a covariate (attained age is thus obtained by summing age at diagnosis and survival time). The link with the international actuarial notation for survival probabilities and hazard is as follows: if the insurance applicant aged x has been diagnosed with cancer at age $x - w$ then

$$
\begin{aligned}
{}_s p_x &= \frac{S(w+s|\text{age at diagnosis} = x-w)}{S(w|\text{age at diagnosis} = x-w)} \\
\mu_{x+s} &= \lambda(w+s|\text{age at diagnosis} = x-w).
\end{aligned}
$$

Relative survival provides a measure of the excess mortality experienced by cancer patients by comparing the mortality in the cancer population with the mortality in the general population. Relative survival models are divided into two types: (i) additive and (ii) multiplicative models. The choice between these models hinges on the underlying assumptions about the relationship between covariates and excess mortality. Multiplicative models assume that covariates proportionally affect the excess mortality rate, making them suitable in contexts where the relative effects of covariates are of primary interest, such as in actuarial studies and general epidemiology. This is consistent with actuarial practices, where mortality rates are typically modeled multiplicatively, although adjustments (such as the correction of the population hazard) may sometimes be additive for specific applications involving multiple causes of death. The latter is precisely the approach followed in biostatistics, where mortality is decomposed into a reference level and its excess due to the disease under consideration. This corresponds to a double decrement model with two causes of death, either cancer or all other causes. The resulting force of mortality is then the sum of these two terms, resulting in an additive specification. Thus, the choice of model should be guided by the specific context of the study and the nature of the data. For a more detailed discussion, see, e.g., Bolard et al. [1], Buckley [3], Dickman et al. [6], Esteve et al. [7], Hakulinen and Tenkanen [10], and Pohar and Stare [16]. While multiplicative specifications are widely accepted within the actuarial community, additive models are commonly used in biostatistics for cancer studies and often provide a better fit to such data. The additive specification is thus adopted in this paper. The hazard $\lambda(t|z)$ at time t since diagnosis for cancer patients with covariate vector z is decomposed into two

additive components: the population hazard based on available patient's characteristics z, denoted as $\lambda_P(t|z)$, and the excess hazard specific for the disease of interest, denoted as $\lambda_E(t|z)$. Formally,

$$\lambda(t|z) = \lambda_P(t|z) + \lambda_E(t|z). \tag{4.1}$$

In (4.1), $\lambda_P(\cdot|z)$ usually corresponds to general population life tables. Here, the covariate vector z corresponds to age and it is the same in $\lambda_P(t|z)$ and $\lambda_E(t|z)$.

Remontet et al. [17] and Fauvernier et al. [8, 9] proposed a flexible parametric model to (i) allow for a flexible modeling of the baseline excess hazard, (ii) account for non-linear and non-proportional effects of covariates and (iii) allow for a flexible interaction between several covariates adopting a multidimensional penalized splines approach. This leads to the specification

$$\ln \lambda_E(t|z) = \sum_{j=1}^{J} g_j(t, z) \tag{4.2}$$

where $g_j(\cdot, \cdot)$ are uni- or multidimensional penalized spline function. This model has the advantage that the splines bring the flexibility needed for modeling the hazard and the penalty terms control this flexibility for smooth estimation. In Soetewey et al. [20], excess hazard was estimated using the `flexrsurv` package in `R` [5], assuming non-linear and non-proportional hazard for age at diagnosis. More precisely, it was based on a model with a spline of the type truncated power basis, with degree 2 and a knot at one year after diagnosis.

As noted earlier, Soetewey et al. [20] suggested determining the waiting period for the RTBF based on the time needed after diagnosis for insurance premiums to return to an acceptable level, as defined by regulatory life tables. Nonetheless, this method requires data spanning up to 30 years post-diagnosis, covering the 10-year standard RTBF period plus the typical 20-year loan duration (to be adapted depending on the context, i.e. in countries with lower/longer time to RTBF and/or for a shorter/longer loan duration). Alternatively, when survival data are available for a shorter period, that is, less than 30 years post-diagnosis, the method relies on the extrapolation of mortality rates, to cover both the standard RTBF period and the typical loan duration. The methodology proposed by Soetewey et al. [20] is challenged by this requirement, as the chosen extrapolation method can significantly influence the resulting waiting period, introducing uncertainty and variability. This is illustrated in the remainder of the present section.

Let us proceed as in Soetewey et al. [20] and determine the waiting period opening the RTBF as the smallest duration after diagnosis so that the expected present value of mortgage insurance benefits gets back to the premium determined according to XK life table. This is represented in Figs. 1 and 2 for a patient diagnosed at the age of 30 and 50, respectively. The left panels are based on the estimation and extrapolation method adopted in Soetewey et al. [20], that is, based on the excess hazard obtained from a model with splines of the type truncated power basis, with degree 2 and a knot at one year after diagnosis. This was implemented in the `R` package `flexrsurv`. The right

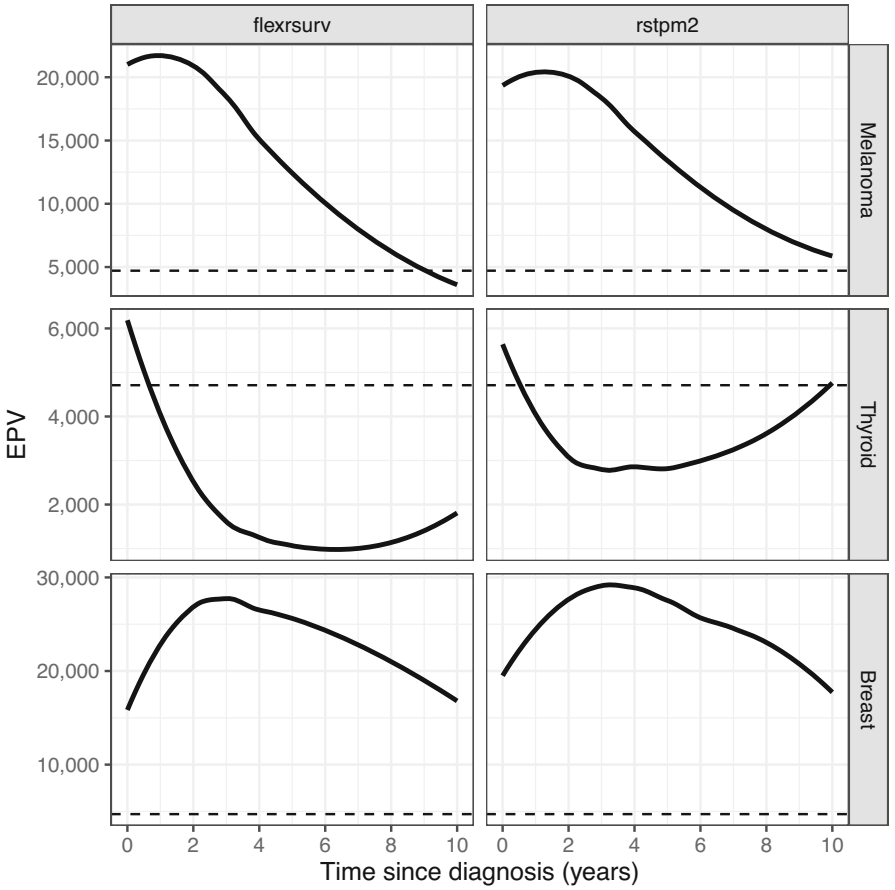

Fig. 1 Expected present value (EPV) of a life insurance contracted by a 30-year-old cancer patient for a period of 20 years with interest of 1% and benefit of 100,000. Horizontal dashed lines correspond to EPV calculated according to XK life table

panels are based on an alternative method implemented in the `R` package `rstpm2` [11, 14, 15, 25]. For this alternative method, we used a flexible parametric survival model with proportional hazards and 3 degrees of freedom for modelling the baseline log-cumulative hazard. These characteristics have been chosen to obtain excess hazards that are as similar as possible to the ones obtained with the `flexrsurv` package, and other scenarios revealed drastically different patterns for the two considered ages at diagnosis. The resulting waiting periods are listed in Table 2. They are obtained by considering that patients become insurable at standard conditions when the EPV reaches the level set by the XK life table. We can see in Table 2 that for melanoma and female breast cancer diagnosed at age 50, the waiting periods determined as in Soetewey et al. [20] are smaller compared to the alternative extrapolation method. For thyroid cancer patients aged 30, the waiting period remains 1 year but EPV exceeds XK level a few years later according to the alternative extrapolation method. For

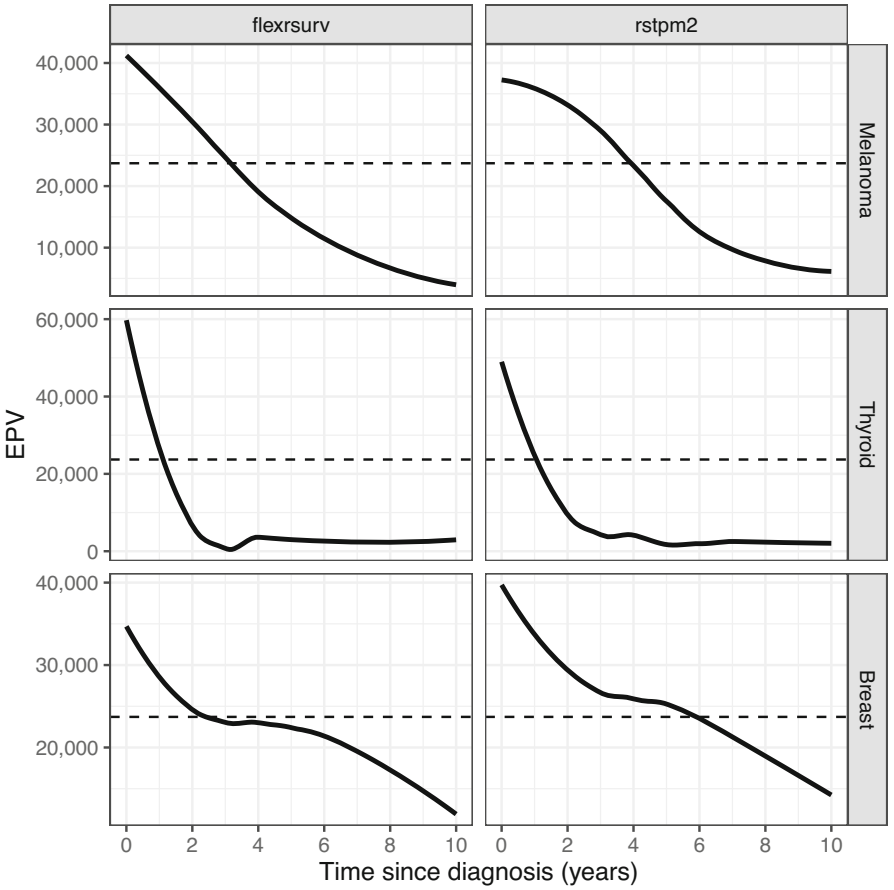

Fig. 2 Expected present value (EPV) of a life insurance contracted by a 50-year-old cancer patient for a period of 20 years with interest of 1% and benefit of 100,000. Horizontal dashed lines correspond to EPV calculated according to XK life table

Table 2 Waiting periods by cancer site and age at diagnosis

Cancer site	Age at diagnosis	Waiting period (in years) Soetewey et al. [20]	Alternative extrapolation method
Melanoma	30	9	> 10
Melanoma	50	3	4
Thyroid	30	1	1*
Thyroid	50	1	1
Breast	30	> 10	> 10
Breast	50	3	6

The star indicates that EPV does not stay below XK level but start to increase a few years after diagnosis

melanoma cancer diagnosed at age 30, the reduced waiting period determined as in Soetewey et al. [20] is contradicted by the alternative extrapolation method.

This example shows that conclusions may rely to a large extent on the extrapolation method, even more so when one considers different model parameters. This is not acceptable in the context of the RTBF. The aim of this paper is to propose a new approach, only using the available data (so without the need to extrapolate mortality rates beyond 10 years), and which is appropriate for any type of cancer. As will be demonstrated in Sect. 5, no extrapolation is necessary in our case since we have follow-up data for more than 10 years post-diagnosis, and our methodology only requires data up to 10 years after diagnosis.

4.2 Conditional relative net survival

Van Ginckel et al. [24] applied a pure biostatistical approach based on an arbitrary cut-off of 0.99 for the conditional net survival to propose reduced waiting periods for breast cancer. This section explains why their apparently sound methodology fails to convince actuaries.

In accordance with actuarial notation, let $q_y = 1 - p_y$ be the one-year death probability at age y (that is, the probability of dying before age $y + 1$ given that the individual is alive at age y). Probabilities, corresponding to general population mortality, are henceforth denoted as p_y^{NIS} and q_y^{NIS} where "NIS" refers to the National Institute of Statistics (Statbel based in Brussels; `www.statbel.fgov.be`). Likewise, denote as $p_{x,w}^{\text{CR}}$ and $q_{x,w}^{\text{CR}}$ these probabilities for an individual of age x who was diagnosed with cancer w years ago, so at age $x - w$. Here, "CR" refers to Cancer Registry established at national level.

The "conditional relative net survival" referred to in Section 4.2.1 of Van Ginckel et al. [24] can be interpreted as the ratio $p_{x,w}^{\text{CR}}/p_y^{\text{NIS}}$. The reduced waiting period is then determined as the smallest w such that $p_{x,w}^{\text{CR}}/p_y^{\text{NIS}} > 0.99$. Their argument is that the resulting w ensures that surviving patients' mortality is very close to the general population one. However, when translated into premium calculation, this rule turns out to produce large increases. Indeed, considering that patients can be covered at standard conditions once they have survived w years after diagnosis, with one-year survival probability

$$p_{x,w}^{\text{CR}} = 0.99 p_y^{\text{NIS}} \tag{4.3}$$

means that

$$q_{x,w}^{\text{CR}} = 1 - p_{x,w}^{\text{CR}} = 1 - 0.99 p_y^{\text{NIS}} = q_y^{\text{NIS}} + 0.01 p_y^{\text{NIS}}.$$

Hence, the one-year death probability (driving the amount of premium for a one-year term insurance) is increased by 1% times the corresponding one-year survival probability. The impact of this rule greatly varies according to age x:

- if $q_y^{\text{NIS}} = 0.001$ then this results in an actual one-year death probability

$$0.001 + 0.01 \times 0.999 = 0.01099$$

which means that the one-year death probability (and hence the yearly term insurance premium) is multiplied by 10, approximately.

- if $q_y^{\text{NIS}} = 0.01$ then this results in an actual one-year death probability

$$0.01 + 0.01 \times 0.99 = 0.0199$$

which means that the one-year death probability (and hence the yearly term insurance premium) is multiplied by 2, approximately.

Considering the typical age range where mortgage insurance is sold, the rule retained by Van Ginckel et al. [24] allows for mortality levels which largely exceed those corresponding to general population.

5 Proposed approach for limited follow-up

Clearly, the rule defining reduced waiting periods for the RTBF must be expressed in terms of premiums. The question about the RTBF centers on evaluating extra claim costs and sharing them among stakeholders in a fair and transparent way. This can only be achieved by computing actual premiums at age x in function of the time w elapsed since diagnosis, and by comparing them to the reference levels XK corresponding to regulatory expected costs.

Let π_0^{XK} be the amount of premium obtained from (2.1) when survival probabilities and death rates correspond to the XK life table. Then, formula (2.1) is used again to obtain the additive increase in mortality compared to the general population level. Precisely, the additive mortality shift γ_y, where the subscript y denotes age at diagnosis, is the unique positive root of the equation

$$\begin{aligned}\pi_0^{\text{XK}} = \sum_{k=0}^{n-1} \exp\left(-\sum_{j=0}^{k-1}\left(\mu_{y+j}^{\text{NIS}} + \gamma_y\right)\right)(1+i)^{-k} c_k \left(\mu_{y+k}^{\text{NIS}} + \gamma_y\right) \\ \times \frac{1 - \exp\left(-\left(\mu_{y+k}^{\text{NIS}} + \gamma_y\right)\right)\frac{1+r}{1+i}}{\mu_{y+k}^{\text{NIS}} + \gamma_y - \ln(1+r) + \ln(1+i)}.\end{aligned} \tag{5.1}$$

The solution is unique because the right-hand side of this equation is increasing in γ_y and the left-hand side is larger than the right-hand side when $\gamma_y = 0$ because the XK life table is conservative. The solution is therefore such that $\gamma_y > 0$.

Following the idea of (4.3), we propose to define the waiting period opening the RTBF as the smallest w such that

$$p_{x,w}^{\text{CR}} = \exp(-\gamma_y) p_y^{\text{NIS}}. \tag{5.2}$$

In this case, we recover a constraint on the conditional observed survival, but with the arbitrary 0.99 level replaced with $\exp(-\gamma_y)$ controlling premium. Following (5.2), the waiting period opening the RTBF is determined as the smallest w such that

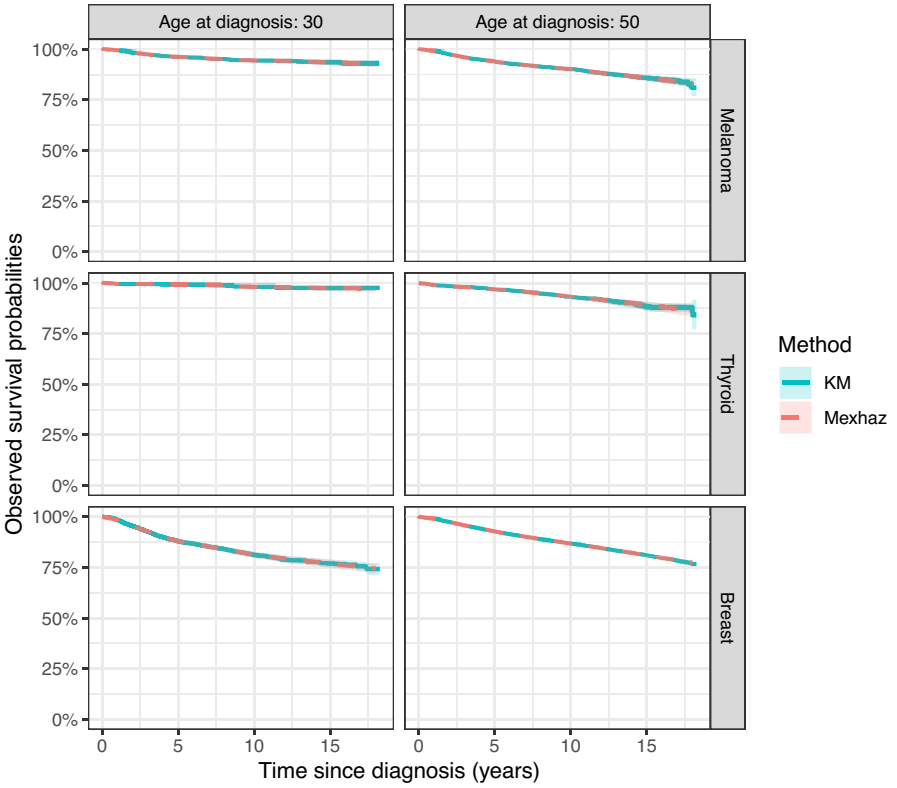

Fig. 3 Survival probabilities (with 95% confidence interval) by cancer site and age at diagnosis. Mexhaz-based method corresponds to the probabilities obtained via a flexible parametric model (dashed line), whereas KM-based method corresponds to the ones obtained based on the non-parametric Kaplan–Meier estimator (solid line)

$p_{x,w}^{\mathrm{CR}}/p_y^{\mathrm{NIS}} > \exp(-\gamma_y)$. To apply this rule, p_y^{NIS} can easily be found within Belgian population life tables, available from Statbel. The calculation of $p_{x,w}^{\mathrm{CR}}$ is explained in Appendix A.

Let us now apply this method to get the length of the waiting period opening the RTBF. To this end, survival probabilities of cancer patients, obtained via a flexible parametric model (using the `mexhaz` R package [4] and based on a baseline hazard specified as the exponential of B-splines of degree 2 with a knot at 2.5 years of follow-up), are first compared with the observed survival probabilities obtained with the non-parametric Kaplan–Meier [12] estimator. The results are displayed in Fig. 3. It can be seen from Fig. 3 that observed survival curves obtained via a flexible parametric model and via the Kaplan–Meier estimator are very similar for all cases under consideration (i.e., for both ages at diagnosis and for all three cancers of interest).

Secondly, conditional one-year observed survival probabilities (denoted $p_{x,w}^{\mathrm{CR}}$) are computed via a flexible parametric model with baseline hazard specified as the exponential of B-splines of degree 2 with a knot at 2.5 years of follow-up (determined after

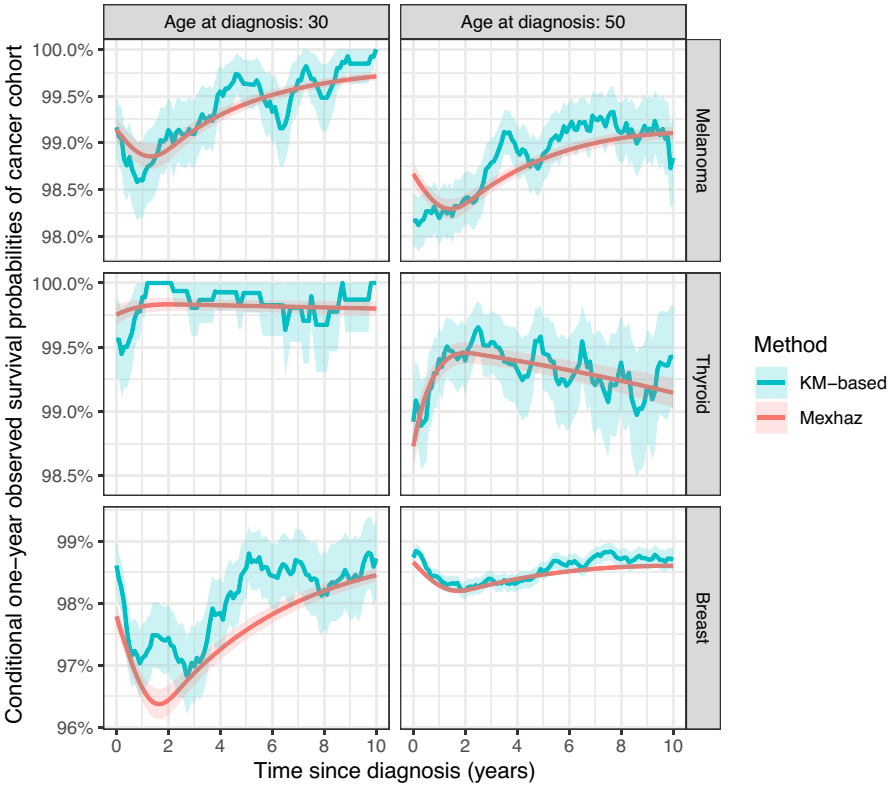

Fig. 4 Conditional one-year survival probabilities (with 95% confidence interval) by cancer site and age at diagnosis. Mexhaz-based method corresponds to the probabilities obtained via a flexible parametric model, $p^{\mathrm{CR}}_{x,w}$, whereas KM-based method corresponds to the ones obtained based on the non-parametric Kaplan–Meier estimator, $p^{\mathrm{KM}}_{x,w}$

having explored different numbers and positions of time knots for the spline basis). More information can be found in Appendix A. $p^{\mathrm{CR}}_{x,w}$ are compared with the conditional one-year observed survival probabilities obtained based on the Kaplan–Meier estimator (henceforth denoted as $p^{\mathrm{KM}}_{x,w}$) in Fig. 4. Here, probabilities $p^{\mathrm{KM}}_{x,w}$ are computed with increments of 0.1 year and are referred as the KM-based method in the remainder of the text since these probabilities are computed based on the Kaplan–Meier estimator. The difference with a standard Kaplan–Meier estimator is that $p^{\mathrm{KM}}_{x,w}$ correspond to conditional one-year observed survival probabilities (instead of simply observed survival probabilities). In practice, $p^{\mathrm{KM}}_{x,w}$ are found by computing one-year survival probabilities using the `survival` R package [23], repeatedly for each subgroup of patients who survived at least $0, 0.1, 0.2, \ldots, 10$ years since diagnosis. The advantage of computing $p^{\mathrm{KM}}_{x,w}$ this way is that the provided confidence intervals are usable, which is not the case if $p^{\mathrm{KM}}_{x,w}$ are computed by dividing the survival probability at a given time by the survival probability one year earlier. The goal of comparing $p^{\mathrm{CR}}_{x,w}$ with a counterpart based on a non-parametric reference such as the Kaplan–Meier estimator

is to demonstrate that results obtained with the proposed approach are trustworthy. Note that, to address concerns about periodicity, probabilities were also calculated using increments of 0.02 year and consistent results were found, leading to the choice of an increment of 0.1 year to balance computational efficiency with accuracy.

Figure 4 shows that conditional one-year survival probabilities obtained via the flexible parametric model follow globally the same trend than the ones obtained via the KM-based method for all scenarios, except for the first year after diagnosis for patients diagnosed with melanoma cancer at age 50. We consider that conditional one-year observed survival probabilities are reasonably well estimated with our approach when compared to a non-parametric reference.

Note that the primary advantage of using a parametric method, such as the flexible parametric model, over non-parametric methods like the Kaplan–Meier estimator lies in its ability to produce more efficient and precise estimates. In this study, the flexible parametric model yields narrower confidence intervals, which is particularly valuable when dealing with smaller sample sizes, where non-parametric methods often result in wider intervals. This is illustrated in Fig. 4, where the confidence intervals are narrower with our approach compared to the non-parametric Kaplan–Meier for a given sample size, highlighting a key advantage of the parametric method over the non-parametric reference. Additionally, the parametric model smooths survival curves, allowing it to capture underlying trends that might be missed by the step-function nature of the Kaplan–Meier estimator. This smoothing is crucial when interpreting conditional survival probabilities, as Kaplan–Meier estimates can fluctuate due to small sample sizes. In summary, while the Kaplan–Meier estimator is a robust and valuable tool, the flexible parametric model offers significant advantages in terms of precision and smoothing, making it a preferred method for estimating conditional one-year survival probabilities in this context.

To determine the waiting period opening the RTBF, conditional one-year survival probabilities obtained via the two approaches, that is, $p_{x,w}^{\text{CR}}$ and $p_{x,w}^{\text{KM}}$, are divided by the conditional one-year survival probabilities in the general population, that is, p_y^{NIS}. Results are displayed in Fig. 5. For the sake of comparison, the waiting period opening the RTBF is determined as the smallest w such that $p_{x,w}^{\text{CR}}/p_y^{\text{NIS}} > \exp(-\gamma_y)$ or such that $p_{x,w}^{\text{KM}}/p_y^{\text{NIS}} > \exp(-\gamma_y)$. Dividing $p_{x,w}^{\text{CR}}$ and $p_{x,w}^{\text{KM}}$ by p_y^{NIS} allows the comparison with the additive correction $\exp(-\gamma_y)$. Here, $\gamma_{30} = 0.0014$ and $\gamma_{50} = 0.0063$. Notice the difference with the threshold of 0.99 set in Van Ginckel et al. [24], as $\exp(-\gamma_{30}) = 0.998601$ and $\exp(-\gamma_{50}) = 0.9937198$. Also note that, for patients aged 30 years at diagnosis, $p_{x,w}^{\text{KM}}$ is actually computed based on patients aged between 25 and 35 years at diagnosis. For patients aged 50 years at diagnosis, $p_{x,w}^{\text{KM}}$ is computed based on patients aged between 45 and 55 years at diagnosis. This is to include more patients and thus have more stable estimates. Indeed, samples of patients diagnosed at exactly 30 and 50 years old have a limited size, in particular for thyroid cancer. Considering patients aged from 25 to 35 and from 45 to 55 instead of patients of exactly 30 and 50 years old does not undermine our analyses, as patients within each age group are very similar in terms of survival.

Results are displayed in Table 3 and Fig. 5. Remember that an increasing ratio is a sign of better prognosis for cancer patients. On the other hand, a decreasing ratio

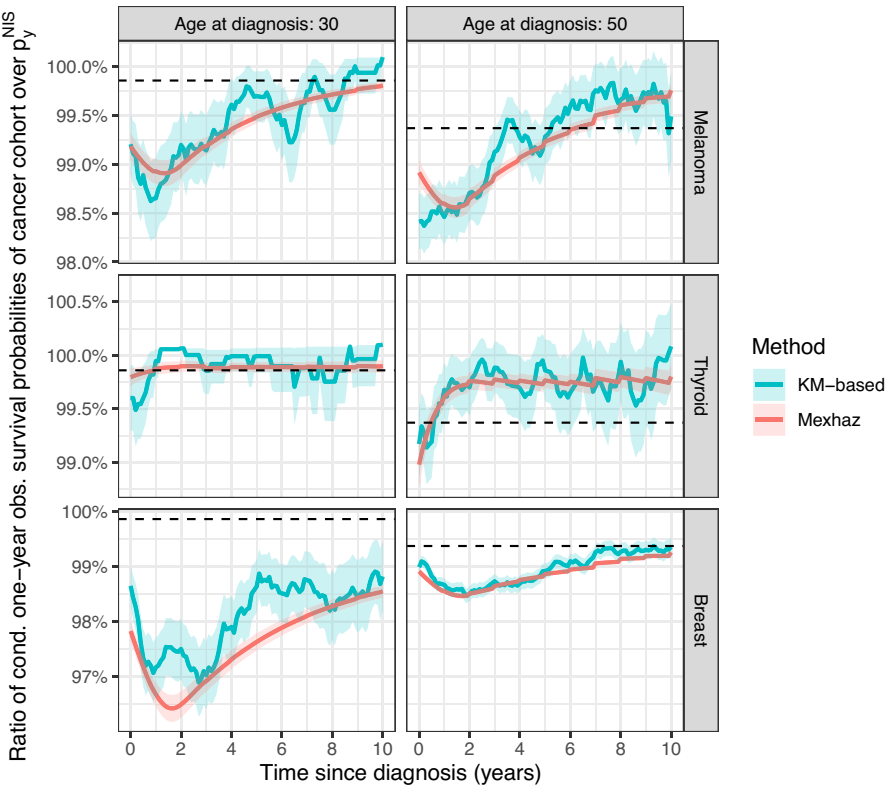

Fig. 5 Ratio of conditional one-year survival probability (with 95% confidence interval) by cancer site and age at diagnosis, together with the additive correction $\exp(-\gamma_y)$ (horizontal dashed lines). Mexhaz-based method corresponds to $p_{x,w}^{\mathrm{CR}}/p_y^{\mathrm{NIS}}$, whereas KM-based method corresponds to $p_{x,w}^{\mathrm{KM}}/p_y^{\mathrm{NIS}}$

Table 3 Waiting periods by cancer site and age at diagnosis computed via our approach and via the KM-based method

Cancer site	Age at diagnosis	Waiting period (in years) Our approach	KM-based
Melanoma	30	> 10	10
Melanoma	50	6	6
Thyroid	30	1	Uncertain
Thyroid	50	1	2
Breast	30	> 10	> 10
Breast	50	> 10	> 10

is a sign that survival for cancer patients declines over the years since diagnosis, so a sign of worse prognosis compared to the general population. Following this, and in order to be as conservative as possible, if the ratio of conditional one-year survival probability reaches the level of the additive correction more than once within the 10-year period after diagnosis, the waiting period is set as the largest time after diagnosis where the ratio of survival probabilities crosses the additive correction level. Notice

also the emergence of small jumps when plotting the ratio of the conditional survival probabilities in Fig. 5 resulting from the division by the one-year survival probabilities p_y^{NIS} in the general population.

From Table 3 and Fig. 5, we can see that waiting periods are below 10 years for melanoma cancer patients aged 50 years at diagnosis, and thyroid cancer patients aged 30 and 50 at diagnosis. Waiting periods obtained via the KM-based method are below 10 years for melanoma and thyroid cancer patients aged 50 at the time of diagnosis. For all other scenarios, waiting periods are equal or above 10 years after diagnosis. When comparing the two approaches, waiting periods are relatively equivalent for all considered subgroups except for thyroid cancer patients diagnosed at 30 years old, who have an uncertain waiting period via the KM-based method (since the ratios of conditional one-year survival probabilities fluctuate around the level set by the additive correction from 0 to 10 years after diagnosis). Moreover, breast cancer patients diagnosed at age 30 have a waiting period above 10 years while it is equal to 10 years according to the KM-based calculation. Recall that, as advocated in Soetewey et al. [20], these waiting periods start at the time of diagnosis, and not at the end of the therapeutic protocol as it is the case with the current legislation.

6 Impact of the stage of the tumor

One could argue that mortality and thus the waiting period opening the RTBF varies between cancer patients diagnosed at different tumor stages. As this information is available in BCR, this section refines the preceding analyses by cancer stage at diagnosis.

Information on both the clinical and pathological staging has been combined to define a final tumor stage. First, clinical staging is an estimate of the extent of the cancer based on results of physical exams, imaging tests, endoscopy exams, biopsies, and for some cancers, the results of other tests, such as blood tests. Second, the pathological staging (also called the surgical stage) is an estimate of the extent of the cancer that is based on the results of pathological examination of the resection piece after surgery. In some cases, the pathological stage is different from the clinical stage, for instance, if the surgery shows the cancer has spread more than was seen on imaging tests. A common practice is to combine these two methods to obtain a so-called combined stage. When the pathological stage is known, it is taken as combined stage, unless there is clinical evidence of metastasis. In case the pathological stage is unknown, the clinical stage is retained. Combining the clinical and pathological stage limits missing values (missing combined stage appears only when both the clinical and pathological stages are missing). This combined stage is considered in this section.

Stages I, II, III and IV were considered. Tumors with an unknown stage at the time of diagnosis, representing 7.5% of all tumors, have been ignored. Number of included cases, number of observed deaths, one-year and 5-year observed survival probabilities (obtained with the non-parametric Kaplan–Meier estimator) by cancer site and stage of the tumor are displayed in Table 4. Given the small number of observations for stages III and IV, these two stages have been combined for the analyses. Furthermore,

Table 4 Number of included cases, number of observed deaths, one-year and 5-year observed survival probabilities (with 95% confidence interval) by cancer site and stage of the tumor

Cancer site	Tumor stage	Number of cases	Number of deaths	1-year survival prob. (95% CI)	5-year survival prob. (95% CI)
Melanoma	I	20,949	1153	0.997 (0.996–0.997)	0.970 (0.968–0.973)
	II	3019	777	0.980 (0.975–0.985)	0.802 (0.786–0.817)
	III	1669	537	0.949 (0.939–0.960)	0.711 (0.688–0.735)
	IV	444	327	0.658 (0.615–0.703)	0.298 (0.257–0.345)
Thyroid	I	8071	311	0.995 (0.994–0.997)	0.979 (0.976–0.982)
	II	946	73	0.989 (0.983–0.996)	0.961 (0.948–0.974)
	III	841	134	0.993 (0.987–0.999)	0.942 (0.926–0.958)
	IV	617	294	0.779 (0.747–0.812)	0.625 (0.587–0.665)
Breast	I	56,039	4813	0.996 (0.995–0.996)	0.965 (0.964–0.967)
	II	37,935	5979	0.992 (0.991–0.993)	0.926 (0.923–0.929)
	III	11,919	3922	0.976 (0.973–0.979)	0.805 (0.798–0.813)
	IV	5639	3839	0.829 (0.819–0.839)	0.390 (0.377–0.404)

Survival probabilities are obtained with the non-parametric Kaplan–Meier estimator

to ensure simplicity and given that cancer patients diagnosed at stages I and II are relatively similar in terms of survival, these two stages have also been combined.

The present section is aimed at studying the appropriateness of stratifying the RTBF according to the stage of the tumor at diagnosis: waiting periods are computed separately for patients diagnosed at stages I–II and at stages III–IV using our proposed approach. This will serve as a comparison with results obtained before, where all stages are included. Note that, as the additive correction $\exp(-\gamma_y)$ depends only on age at diagnosis, it differs between patients diagnosed at 30 and 50 years old, but it is the same for all stages and it remains the same when including all stages. Notice that the non-parametric Kaplan–Meier reference is no longer used because stratifying by stage reduces drastically the number of observations, in particular for stages III and IV. This rises the issue of the accuracy of the Kaplan–Meier estimator, and therefore reduces its usefulness in the context of the RTBF.

Results of the stratification by stage at diagnosis are displayed in Table 5 and Fig. 6. Waiting period is lower for patients diagnosed at stages I–II compared to patients diagnosed at stages III–IV for all scenarios. In particular, compared to patients diagnosed at all stages, when including only patients diagnosed at stages I–II, waiting periods are reduced from 6 to 4 years for melanoma cancer patients aged 50, reduced from 1 to 0 year for thyroid cancer patients aged 50, and reduced from more than 10 years to 7 years for female breast cancer patients aged 50. For melanoma cancer patients aged 30, thyroid cancer patients aged 30 and breast cancer patients aged 30, waiting periods remain the same whether it is calculated by stage or for all stages combined. This shows that, for the three cancer sites considered, stratifying the analyses according to the stage has no impact on the waiting periods for patients diagnosed at the age of 30, but has an impact for patients diagnosed at the age of 50. This can be partly explained by the fact that, among patients diagnosed at a young age, a small proportion

Table 5 Comparison of waiting periods by cancer site and age at diagnosis resulting from our approach and from Soetewey et al. [20]

Cancer site	Age at diagnosis	Proposed approach Stages I–II	Stages III–IV	All stages	Soetewey et al. [20] All stages
Melanoma	30	> 10	> 10	> 10	9
Melanoma	50	4	> 10	6	3
Thyroid	30	1	> 10	1	1
Thyroid	50	0	> 10	1	1
Breast	30	> 10	> 10	> 10	NA
Breast	50	7	> 10	> 10	NA

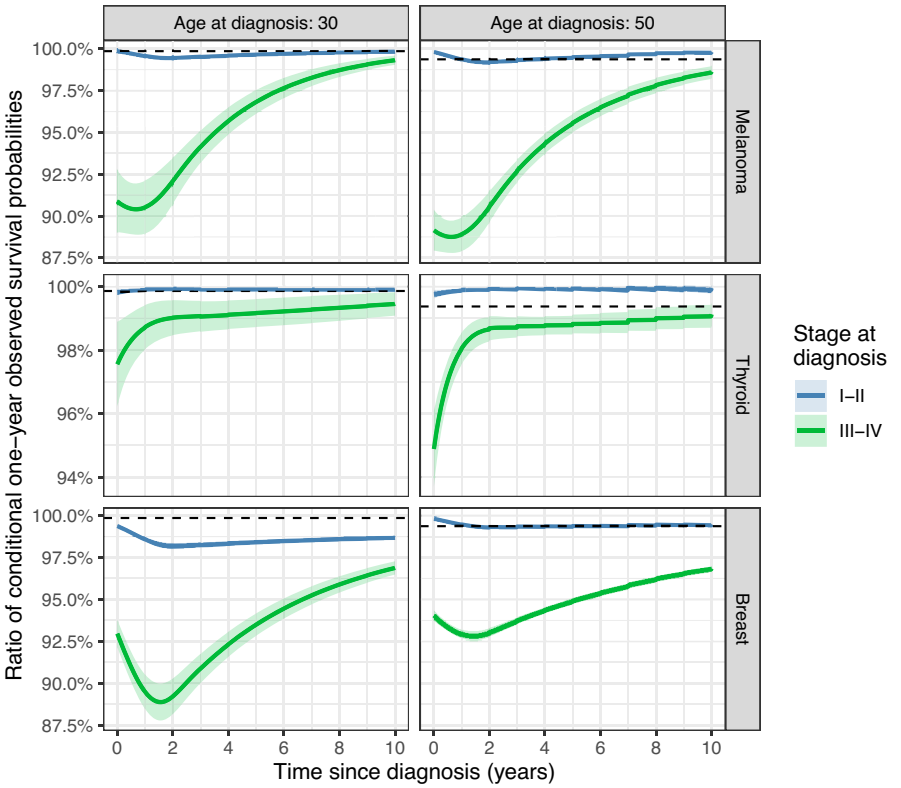

Fig. 6 Ratio of conditional one-year survival probability obtained via the proposed approach (i.e., $p_{x,w}^{CR}/p_y^{NIS}$) by cancer site, age and stage at diagnosis. Horizontal dashed lines correspond to the additive correction $\exp(-\gamma_y)$

is diagnosed at stages III–IV. For instance, only 8.36% of patients aged 30 or below at the time of diagnosis are diagnosed at stages III–IV. Furthermore, we observe that the waiting period is above 10 years for patients diagnosed at stages III–IV for all scenarios.

Notice that 95% confidence intervals for stages I–II (Fig. 6) are narrower than when all stages are considered (Fig. 5), although the sample size is smaller when including only patients diagnosed at stages I–II. This is explained by the fact that, for the same sample size, standard errors for probabilities closer to 0% or 100% are smaller. Therefore, even though the sample size is smaller for stages I–II than for all stages combined, confidence intervals are narrower because probabilities are closer to 100% for this subgroup of patients. Also notice that confidence intervals do not widen with time since diagnosis, contrarily to what would be expected given that the sample size decreases with time since diagnosis. The following elements explain this phenomenon. We only consider age at diagnosis 20–69 years. Survival is high for this age range and the cancer types considered, so 10 years after diagnosis will not yet be long enough to see a clear increase in the length of the confidence intervals. And again, when survival probabilities approach 100%, the confidence intervals become smaller for a given number of observations. A similar pattern for the conditional net survival has been found in Van Ginckel et al. [24] for female breast cancer. Calculations of the confidence intervals are further explained in Appendix A.

7 Discussion

The methodology proposed in this paper addresses the issue of limited follow-up by focusing on using available data without the need for extrapolation beyond 10 years post-diagnosis. This approach is particularly advantageous for cases where follow-up is shorter than the required 30 years post-diagnosis (i.e., 10-year standard RTBF period plus the typical 20-year loan duration), as it avoids the uncertainties inherent in long-term extrapolation of mortality rates. Importantly, this method does not require extrapolation beyond 10 years whatever the type of cancer. By avoiding the need for extrapolation, this methodology ensures a more consistent and reliable determination of the waiting period for RTBF across various cancer types. In our study, no extrapolation was needed, as our method requires follow-up data only up to 10 years after diagnosis, which aligns with the data available.

To sum up, let us compare waiting periods obtained with our approach with results obtained according to the method proposed by Soetewey et al. [20], which are based on the time after diagnosis when the expected present value of a standard mortgage insurance reaches the same level than the one based on XK life table. A summary is displayed in Table 5. We can see that waiting periods are sensibly the same for thyroid cancer patients across all methods, while they are slightly higher when estimated via the approach proposed in this paper for melanoma cancer patients. Note that no comparison is made for breast cancer, as this cancer site was not considered in Soetewey et al. [20].

Results in Table 5 are in line with the reduced waiting periods specified in the Belgian legislation. Furthermore, results are also in line with the AERAS convention

(i.e., the reference grid used in France), which stipulates that the RTBF is maximum 6 years after the end of the therapeutic protocol for melanoma and thyroid cancers.

Nonetheless, an important difference is that in this paper, all waiting periods opening the RTBF are based on the time since diagnosis, rather than on the time since the end of the therapeutic protocol as currently implemented in the Belgian and French reference grids. As duration of cancer treatments are unpredictable and heterogeneous (even within the same cancer site and stage), a RTBF based on the date of diagnosis rather than based on the treatment end date will benefit both patients and insurers. Indeed, patients will know exactly when they can expect to benefit from this RTBF, and insurers will face less uncertainties (as the date of diagnosis is known and fixed, contrarily to the treatment end date which is difficult to establish and may change over time depending on the patient's health status) and less prone to debates (about, for instance, what is considered as treatment or not).

Although data used in the analyses cover a relatively long period of time (year of diagnosis ranges from 2004 to 2020) with diagnostic criteria and methods that have evolved and improved over that period, calendar time has not been included for two main reasons. First, the limited number of cases available (in particular since the focus is on young adults) prevents another division between different cohorts. Second, given that medicine and treatments progress with time, survival of cancer patients also improve with time. Thus, the resulting potential bias of omitting a cohort effect appears to be conservative, as the actual time for the patients to reach a survival comparable to that of the general population will decrease with improving treatments. In addition to that, population data are used whereas outstanding balance insurance applicants belong to the upper socio-economic class who usually have better prognosis, and individuals who contract a home or professional loan are generally in good health as individuals with poor health are unlikely to embark on such a project. These selection effects imply that analyses conducted in the present paper are conservative in many respects.

One could argue that waiting periods are expected to be shorter for patients diagnosed at stages III–IV compared to patients diagnosed at stages I–II, as we would expect when comparing patients diagnosed with pancreatic and breast cancer. The idea behind this reasoning is that the worse the prognosis, the quicker the patients die after diagnosis and thus the quicker only the survivors remain. Results of the stratification by stage show that it is not the case. The following arguments explain it. Statistical cure in the case of female breast cancer is not yet achieved within 15 years after diagnosis (except for stage I), while it is achieved for pancreatic cancer at around 5 years after diagnosis. Indeed, for female breast cancer, excess hazard is relatively constant and non negligible even after many years after diagnosis, with late recurrences occurring up to 20 years after diagnosis. On the contrary, for aggressive cancers, excess hazard is much less constant over the years after diagnosis, and in the case of pancreatic cancer it becomes negligible around 5 years after diagnosis. Given the difference in excess mortality between breast and pancreatic cancer, it is reasonable to expect waiting periods to be shorter for pancreatic than for breast cancer. Although melanoma and thyroid cancers are nowhere near as aggressive as pancreatic cancer, the trend of the excess hazard for these two cancers is closer to pancreatic than to breast cancer, that is, excess hazard is not constant over the years after diagnosis, it

becomes negligible only after some years after diagnosis and late recurrences are rare. This explains the shorter waiting periods for melanoma and thyroid cancers compared to female breast cancer. The same reasoning can be applied to the comparison of the waiting periods between stages of the tumor. One could expect that the more advanced the stage, the more quickly only the survivors remain and thus the shorter the waiting period. This holds only if statistical cure is reached at a given time after diagnosis (and in particular within 10 years after diagnosis to argue for a reduced waiting period opening the RTBF). For female breast cancer, the excess hazard for stages III–IV is higher than for stages I–II up to 15 years after diagnosis, resulting in waiting periods that are not shorter for stages III–IV compared to stages I–II.

Results obtained in the present study focus on melanoma, thyroid and female breast cancer patients for illustrative purposes. The approach developed in this paper can be applied to other cancer types or diseases. However, as just discussed, it cannot be used in case of late recurrences nor to chronic diseases to argue a shorter waiting period. For some cancers with late recurrences such as breast cancer, the waiting period resulting from our proposed approach when including patients diagnosed at all stages of the tumor is (much) longer than if it was proposed only to patients diagnosed at stages I–II. For melanoma and thyroid cancers, the waiting period resulting from our proposed approach when including patients diagnosed at all stages of the tumor is relatively similar than if it was proposed only to patients diagnosed at stages I–II. This demonstrates that, besides the fact that computing the waiting period should be done by stage for female breast cancer while it is not compulsory for melanoma and thyroid cancers, cancer is not one disease, but a family of many diverse diseases with different outcomes. Therefore, the proposed method should be applied on a case-by-case basis, that is, cancer by cancer. This is left for future research.

As mentioned in Sect. 1, the method proposed by Soetewey et al. [20] remains actuarially sound if the length of the follow-up is long enough. It could be argued, however, that even when registry data have a sufficiently long follow-up period, the method proposed in this paper would still be preferable since a long follow-up means that some patients have been diagnosed a long time ago, and are thus not treated as well as nowadays. This argument is all the more valid the longer the follow-up time, as the longer the follow-up, the greater the potential increase in treatment efficacy between the beginning and end of the follow-up period.

The proposed approach can obviously be applied in other countries by replacing the databases by the appropriate ones. Moreover, other cancer sites and other diseases which qualify for the RTBF (e.g., HIV, some types of hepatitis and leukemia) are left for future research. This would undeniably be useful to improve the reference grids in Belgium and other countries, and ultimately, to improve access to such insurance products for other types of surviving patients.

There are several considerations and complexities surrounding the RTBF and its application in insurance contexts. We highlight seven of them. First, it is crucial to consider the impact on the insurance portfolio and pricing. The European Commission's guidelines on the RTBF, while aiming to protect cancer survivors and other individuals with prior health conditions, often leave room for interpretation at the member state level. This has led to significant variations in the implementation of the RTBF, with some countries reducing waiting periods or extending the RTBF to

include conditions beyond cancer, such as hepatitis C, diabetes or cystic fibrosis. Such deviations from the EU guidelines can affect the calculation of individual risk-specific premiums. Specifically, when the RTBF limits the use of individual health histories, insurers may be forced to distribute the associated risk across the entire insured pool. This redistribution can lead to higher overall premiums for all policyholders, as the risk of insuring individuals with past health conditions is spread more broadly. The magnitude of this impact depends on the proportion of the insured population affected and the extent of the deviations from the standard risk assessment models. To mitigate these effects, the life insurance industry may need to adapt by refining risk assessment techniques, employing more sophisticated data analytics, and considering alternative pricing models that can balance fairness with financial sustainability.

Second, it is also important to note that risk-adequate underwriting in life insurance is fundamentally dependent on the evidence-based assessment of long-term health risks, particularly when there is asymmetry in the information available to insurers and applicants. Proper risk-based underwriting practices are crucial not only for ensuring the financial stability of the insurance pool but also for extending coverage to long-term survivors of chronic diseases under the RTBF. An alternative could be to group all insurance applicants benefiting from RTBF (remember that Belgian law imposes the disease to be declared to the insurer) in a dedicated pool and to distribute the losses generated by excess mortality over the entire insurance market, in proportion of each insurer's market share. A similar mechanism is applied in Belgium to fire insurance, in order to maintain solidarity in case of natural disasters.

Third, while the assumption that mortality for cancer patients temporarily peaks after diagnosis before stabilizing to a level comparable to the general population is generally valid for many cancer types, particularly those diagnosed at an early stage, it does not universally apply. Certain cancers, such as breast cancer, ovarian cancer and Hodgkin lymphoma, can exhibit significant excess mortality and morbidity even after long periods of being disease-free. In these cases, the mortality patterns may be more erratic and non-linear, challenging the assumption that a standard waiting period will suffice to equalize the EPV of mortgage insurance benefits with the premiums based on the Belgian regulatory life table. For these types of cancers, applying the RTBF could result in inadequate standard rates, as the long-term risk remains higher than anticipated. To address these cases, it may be necessary to extend the waiting period or apply a differentiated approach that takes into account the specific mortality trajectory of the cancer type. This could involve additional stratification by cancer type or stage at diagnosis, as well as more personalized risk assessments to ensure that insurance premiums reflect the actual long-term risks more accurately. These adjustments would help maintain the actuarial fairness of the insurance offerings while still aligning with the principles of the RTBF.

Fourth, the presence of comorbidities poses a significant challenge in the risk assessment of cancer survivors, as these individuals are more likely to develop conditions such as cardiovascular diseases, diabetes, hypertension, obesity, depression, anxiety, and immune system disorders. These comorbidities can lead to an increased rate of cancer recurrence and elevated mortality from other causes, complicating the application of standard risk models under the RTBF framework. This is the reason why Belgian law requires insurance applicants to declare previous diseases to the insurer (even if

RTBF prohibits using this information in pricing), so that the effect of comorbidities can be assessed. To effectively handle this risk, insurance companies may need to implement more comprehensive underwriting practices that account for the potential impact of comorbidities on long-term mortality. This could involve enhanced screening processes that consider a survivor's overall health profile, including the likelihood of developing comorbid conditions based on the type and stage of cancer, treatment methods, and individual patient characteristics. Additionally, insurers might develop tailored premium structures or offer specialized products that reflect the heightened risk associated with comorbidities in cancer survivors. By doing so, insurers can balance the goals of the RTBF with the need to maintain accurate risk assessment and ensure the financial viability of their insurance portfolios.

Fifth, while our approach focuses on a parallel shift in the force of mortality curve based on the age at diagnosis, we acknowledge that alternative relational models incorporating additional covariates and non-linear specifications may offer potential benefits. However, the utility of such models is contingent upon data availability and the specific context of the analysis. Frailty models, which introduce random effects to capture heterogeneity within a cohort, could theoretically enhance the stratification of excess mortality risk by accounting for individual-level variability [13]. Yet, given that our model already includes critical factors such as cancer stage, further stratification based on more detailed tumor characteristics (e.g., morphology, genetic profiling) may be a more pragmatic direction for future research. The exploration of these alternative modeling approaches represents an area for further investigation, particularly if more granular data become available.

Sixth, multistate models offer a dynamic framework for estimating the RTBF by modeling transitions from an illness state back to a healthy state. This approach can effectively capture the gradual reversion of mortality rates to standard levels, making it suitable for determining RTBF waiting periods. However, these models require detailed transition data, which may not always be available, and their complexity can complicate implementation and interpretation [22]. In contrast, the model proposed in this paper provides a more straightforward method by adjusting mortality rates through a parallel shift based on age at diagnosis. This approach is easier to apply and interpret, though it may be less flexible than multistate models. The choice between these approaches depends on data availability, model complexity, and practical considerations.

Seventh, the choice between period and cohort approaches in mortality modeling is crucial for life insurance and pensions pricing. In this study, we employed a cohort approach to mortality modeling, which uses the full follow-up data of patients diagnosed in the past. This method provides a detailed and precise survival analysis, leveraging all available information to produce robust estimates. While a period approach offers more "up-to-date" survival estimates that may better reflect future outcomes for patients diagnosed today [2], it does so at the expense of using less data, leading to larger standard errors and less precision in determining waiting periods. Although a period analysis could provide estimates closer to future observations, the cohort approach was chosen for its ability to deliver more accurate and reliable results based on comprehensive historical data, ensuring a solid foundation for the estimation of waiting periods under the RTBF framework.

Addressing these complexities and considerations is essential for effectively applying the RTBF in insurance contexts, ensuring both fairness and financial stability within the industry.

A Conditional one-year observed survival probability

A.1 Observed survival and hazard

From the relation between cumulative hazard for all cause death, Λ, follows the observed survival, OS:

$$OS(t) = \exp\left(-\Lambda(t)\right) = \exp\left(-\int_0^t \lambda(u)du\right), \tag{A.1}$$

with $\lambda(u)$ the hazard at time u.

Conditional one-year OS at time t can be obtained from Eq. (A.1) by integrating only over the interval $[t, t+1]$:

$$OS(t, t+1) = \exp\left(-\int_t^{t+1} \lambda(u)du\right). \tag{A.2}$$

To calculate this integral in practice, numerical integration can be applied on a set of time values, say with a step of 0.01 year.

A.2 Flexible parametric model for the hazard

The hazard as a continuous function of survival time was obtained from a flexible parametric model (fpm) using the `mexhaz` function from the R package `mexhaz` [4].

A.3 Predicted observed survival

The observed survival at a given time t, can be obtained from numerical integration of Eq. (A.1):

$$OS(t) = \exp\left(-\sum_{i=0}^{N_t-1} \lambda(t_i)(t_{i+1} - t_i)\right) = \exp\left(-\sum_{i=0}^{N_t-1} \lambda(t_i)\Delta t\right) = \exp\left(-\tilde{\Lambda}(t)\right), \tag{A.3}$$

when the $[0, t]$ interval is split in N_t intervals of width Δt ($t_0 = 0, t_1 = \Delta t, t_2 = 2\Delta t, \ldots, t_{N_t} = t$).

To obtain a curve of the observed survival at a set of time values (say from 0 to 10 years in steps of $\Delta t = 0.01$ year, so 1000 data points), the vector of the corresponding cumulative hazards, $\tilde{\mathbf{\Lambda}}$, calculated via numerical integration is needed:

$$\mathbf{OS} = \exp(-\tilde{\mathbf{\Lambda}}). \tag{A.4}$$

The cumulative hazard vector can be calculated from the estimated regression coefficients via matrix multiplication. Let $\mathbf{X}$ be the design matrix (N_t lines, each line corresponds with a time value) for the needed linear combinations of estimated regression coefficients, $\boldsymbol{\beta}$, at the $\log(\lambda(t))$ scale. The estimated $\mathbf{log(\lambda)}$ vector and its covariance matrix at each time points equals:

$$\mathbf{log(\lambda)} = \mathbf{X}\boldsymbol{\beta} \tag{A.5}$$

$$\sum_{\mathbf{log(\lambda)}} = \mathbf{X}\sum_{\boldsymbol{\beta}}\mathbf{X}^T. \tag{A.6}$$

So:

$$\boldsymbol{\lambda} = \exp(\mathbf{X}\boldsymbol{\beta}) \tag{A.7}$$

$$\sum_{\boldsymbol{\lambda}} = \mathbf{J}_{\boldsymbol{\lambda}}\sum_{\boldsymbol{\beta}}\mathbf{J}_{\boldsymbol{\lambda}}{}^T, \tag{A.8}$$

with $\mathbf{J}_{\boldsymbol{\lambda}}$ the Jacobian matrix $\mathbf{J}_{\boldsymbol{\lambda}} = diag(\exp(\mathbf{X}\boldsymbol{\beta}))$.

The cumulative hazard at all time points is easily obtained by multiplying with a upper triangular matrix $\boldsymbol{T}$ (with 1's on the diagonal):

$$\tilde{\boldsymbol{\Lambda}} = \Delta t \cdot \boldsymbol{\lambda}^T\mathbf{T} \tag{A.9}$$

$$\sum_{\tilde{\boldsymbol{\Lambda}}} = (\Delta t)^2\,\mathbf{T}^T\sum_{\boldsymbol{\lambda}}\mathbf{T}. \tag{A.10}$$

The variances of the cumulative hazards are the diagonal elements of covariance matrix, which allows to construct an asymptotic normal confidence interval (CI).

From Eq. (A.4), it follows:

$$\mathbf{OS} = \exp\left(-\tilde{\boldsymbol{\Lambda}}\right) \tag{A.11}$$

$$\sum_{\mathbf{OS}} = \mathbf{J}_{\mathbf{OS}}\sum_{\tilde{\boldsymbol{\Lambda}}}\mathbf{J}_{\mathbf{OS}}{}^T. \tag{A.12}$$

An asymptotic CI on the obtained survival can be obtained by transforming the CI on the cumulative hazard.

A.4 Predicted conditional one-year observed survival

To obtain the conditional one-year observed survival at each time point t_i, the cumulative hazard over only the next 1 year interval $[t_i, t_i + 1]$ is needed. This can be achieved by creating an upper triangular matrix, $\mathbf{T_c}$, for which the number of 1's in each row is limited up to the next $\frac{1}{\Delta t}$ columns. Take as an example $\Delta t = 0.2$ and consider the

first six lines and the first 10 columns:

$$\mathbf{T_c} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{A.13}$$

The cumulative hazard for the conditional one-year OS becomes:

$$\tilde{\mathbf{\Lambda}}_c = \Delta t \cdot \mathbf{T}_c \boldsymbol{\lambda} \tag{A.14}$$

$$\sum_{\tilde{\mathbf{\Lambda}}} = (\Delta t)^2 \, \mathbf{T}_c \sum_{\boldsymbol{\lambda}} \mathbf{T}_c^T. \tag{A.15}$$

The rest is similar to the observed survival in the previous subsection.

Acknowledgements The three first authors thank UCLouvain for funding this research project, and gratefully acknowledge the Belgian Cancer Registry for providing access to the data and for research assistance. They also gratefully acknowledge funding from the FWO and F.R.S.-FNRS under the Excellence of Science (EOS) program, project EOS 40007517. The authors thank the staff of the Belgian Cancer Registry and all physicians, pathologists and data managers involved in Cancer Registration in Belgium for their dedicated data collection. The authors also thank the two anonymous referees for comments that have been very helpful for revising a previous version of the present work.

Author Contributions All authors contributed to the study conception, design, methodology, data analysis and interpretation. The first draft of the manuscript was written by AS, CL and MD and all authors commented on previous versions of the manuscript. All authors revised the draft and approved the final version of the manuscript.

Funding This research project was supported by UCLouvain and funding from the FWO and F.R.S.-FNRS under the Excellence of Science (EOS) program, project EOS 40007517.

Data Availability The datasets generated and/or analysed during the current study are not publicly available due to privacy reasons but are available from the corresponding author on reasonable request. The pseudonymized data can be provided within the secured environment of the Belgian Cancer Registry according to its regulations, and only upon approval by the Information Security Committee.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Bolard P, Quantin C, Esteve J, Faivre J, Abrahamowicz M (2001) Modelling time-dependent hazard ratios in relative survival: application to colon cancer. J Clin Epidemiol 54(10):986–996
2. Brenner H, Söderman B, Hakulinen T (2002) Use of period analysis for providing more up-to-date estimates of long-term survival rates: empirical evaluation among 370 000 cancer patients in Finland. Int J Epidemiol 31(2):456–462
3. Buckley J (1984) Additive and multiplicative models for relative survival rates. Biometrics 40:51–62

4. Charvat H, Belot A (2021) mexhaz: an R package for fitting flexible hazard-based regression models for overall and excess mortality with a random effect. J Stat Softw 98(14):1–36
5. Clerc-Urmès I, Grzebyk M (2023) flexrsurv: flexible relative survival analysis. R package version 2:17
6. Dickman PW, Sloggett A, Hills M, Hakulinen T (2004) Regression models for relative survival. Stat Med 23(1):51–64
7. Esteve J, Benhamou E, Croasdale M, Raymond L (1990) Relative survival and the estimation of net survival: elements for further discussion. Stat Med 9(5):529–538
8. Fauvernier M, Remontet L, Uhry Z, Bossard N, Roche L (2019) survPen: an R package for hazard and excess hazard modelling with multidimensional penalized splines. J Open Source Softw 4(40):1434
9. Fauvernier M, Roche L, Uhry Z, Tron L, Bossard N, Remontet L, and the Challenges in the Estimation of Net Survival Working Survival Group C (2019b) Multi-dimensional penalized hazard model with continuous covariates: applications for studying trends and social inequalities in cancer survival. J R Stat Soc Ser C (Appl Stat)
10. Hakulinen T, Tenkanen L (1987) Regression analysis of relative survival rates. J R Stat Soc Ser C (Appl Stat) 36(3):309–317
11. Jakobsen LH, Andersson TM-L, Biccler JL, Poulsen LØ, Severinsen MT, El-Galaly TC, Bøgsted M (2020) On estimating the time to statistical cure. BMC Med Res Methodol 20:1–13
12. Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. J Am Stat Assoc 53(282):457–481
13. Legrand C (2021) Advanced survival models. CRC Press, Boca Raton
14. Liu X-R, Pawitan Y, Clements MS (2017) Generalized survival models for correlated time-to-event data. Stat Med 36(29):4743–4762
15. Liu X-R, Pawitan Y, Clements M (2018) Parametric and penalized generalized survival models. Stat Methods Med Res 27(5):1531–1546
16. Pohar M, Stare J (2006) Relative survival analysis in R. Comput Methods Progr Biomed 81(3):272–278
17. Remontet L, Uhry Z, Bossard N, Iwaz J, Belot A, Danieli C, Charvat H, Roche L, Li CWS (2019) Flexible and structured survival model for a simultaneous estimation of non-linear and non-proportional effects and complex interactions between continuous variables: Performance of this multidimensional penalized spline approach in net survival trend analysis. Stat Methods Med Res 28(8):2368–2384
18. Scocca G, Meunier F (2020) A right to be forgotten for cancer survivors: a legal development expected to reflect the medical progress in the fight against cancer. J Cancer Policy 25:100246
19. Scocca G, Meunier F (2022) Towards an EU legislation on the right to be forgotten to access to financial services for cancer survivors. Eur J Cancer 162:133–137
20. Soetewey A, Legrand C, Denuit M, Silversmit G (2021) Waiting period from diagnosis for mortgage insurance issued to cancer survivors. Eur Actuar J 11(1):135–160
21. Soetewey A, Legrand C, Denuit M, Silversmit G (2022) Semi-markov modeling for cancer insurance. Eur Actuar J 12(2):813–837
22. Soetewey A, Legrand C, Denuit M, Silversmit G (2024) Health indices for disease incidence risk and duration in the semi-markov setting. Technical report, Université catholique de Louvain, Institute of Statistics, Biostatistics and Actuarial Sciences
23. Therneau TM, Grambsch PM (2000) Modeling survival data: extending the Cox model. Springer, New York
24. Van Ginckel A, Silversmit G, Van Gool B, Van Damme N, Jonckheer P (2022) The right to be forgotten in breast cancer: new propositions. Belgian Health Care Knowledge Centre (KCE), KCE Reports 351
25. Zhan Y, Liu X-R, Reynolds CA, Pedersen NL, Hägg S, Clements MS (2018) Leukocyte telomere length and all-cause mortality: a between-within twin study with time-dependent effects using generalized survival models. Am J Epidemiol 187(10):2186–2191