Year : 2023

Predicting cancer incidence in Switzerland

Trächsel Bastien

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Unisanté

Predicting cancer incidence in Switzerland

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Bastien Trächsel

Maîtrise universitaire en Statistique, Université de Genève, Suisse

Jury

Prof. Lucia Mazzolai, Présidente
Dre Isabella Locatelli, Directrice de thèse
Dr Jean-Luc Bulliard, Co-directeur de thèse
Prof. Ruth Etzioni, Experte
Prof. Joachim Marti, Expert

Lausanne
(2023)

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Ecole Doctorale
Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof.	David	**Nanchen**
Directeur·trice de thèse	Madame	Dre	Isabella	**Locatelli**
Co-directeur·trice	Monsieur	Dr	Jean-Luc	**Bulliard**
Expert·e·s	Madame	Prof.	Ruth	**Etzioni**
	Monsieur	Prof.	Joachim	**Marti**

le Conseil de Faculté autorise l'impression de la thèse de

Bastien Gilbert Trächsel

Maîtrise universitaire en Statistique, Université de Genève, Suisse

intitulée

Predicting cancer incidence in Switzerland

Lausanne, le 27 juillet 2023

pour le Doyen
de la Faculté de biologie et de médecine

Prof. David Nanchen

Summary

Projecting the future of cancer incidence in a country is an important task for planning future cancer interventions and research and for allocating economic resources. This is a complex exercise, however, as is any attempt to anticipate the future. Applying leave-future-out cross-validation to data from three Swiss cancer registries (Vaud, Geneva, and Neuchâtel) and the period 1982-2016, we compared the predictive performance of a large number of models used in the cancer prediction literature: widely used age-period-cohort (APC) models and their Bayesian counterparts (BAPC), classical generalized linear models (GLM), autoregressive integrated moving average (ARIMA) models, and linear models (LM). Perhaps surprisingly, we found that the simpler a model is, the better it performs in predicting future cancer incidence, in line with the famous Occam's razor principle, which recommends looking for explanations constructed with the smallest possible set of elements. Models simply extrapolating past tendencies (ARIMA, LM) outperformed models seeking to estimate and then project underlying effects (GLM, APC, BAPC). Among the first, models relying on few parameters (e.g. low-order ARIMA) outperformed more complex higher-order models that closely fit observed data, as well as methods based on the well-known AIC selection criterion.

The best model in our comparative study, an ARIMA(2,1,1), was applied to predict cancer incidence in Switzerland until 2025, anticipating a substantial stabilization of the risk of developing cancer for the next few years. Combining this trend with the demographic projections of the Swiss Federal Statistical Office, however, we anticipated a substantial increase in the annual number of new cancer cases, entirely due to demographic changes. This increase was estimated at +18% for men and +11% for women, with increases ranging from 4.15% for thyroid in men to 26% for bladder in men.

Estimating (and predicting) trends in cancer incidence over time can be confounded by changes in cancer detection, such as but not limited to: the introduction or modification of screening programs, the use of different screening tools, and incidental detection. In the third part of this thesis, we proposed a model capable of adjusting for these changes and thus estimating the true underlying trend in cancer incidence.

Résumé

La projection de l'avenir de l'incidence du cancer dans un pays est une tâche importante pour planifier les futures interventions et recherches sur le cancer et pour optimiser l'allocation des ressources. Il s'agit toutefois d'un exercice complexe, comme toute tentative d'anticiper l'avenir. En appliquant une validation croisée aux données de trois registres suisses du cancer (Vaud, Genève et Neuchâtel) pour la période 1982-2016, nous avons comparé la performance prédictive d'un grand nombre de modèles utilisés dans la littérature : les modèles âge-période-cohorte (APC) et leurs équivalents bayésiens (BAPC), les modèles linéaires généralisés (GLM), les modèles autorégressifs intégrés à moyenne mobile (ARIMA) et les modèles linéaires (LM). De manière peut-être surprenante, nous avons constaté que plus un modèle est simple, plus il est performant dans la prédiction, conformément au célèbre principe du rasoir d'Occam, qui veut que la solution la plus simple soit préférée. Les modèles qui se contentent d'extrapoler les tendances passées (ARIMA, LM) sont plus performants que ceux qui tentent d'estimer puis de projeter des effets sous-jacents (GLM, APC, BAPC). Le modèle le plus performant a été L'ARIMA (2,1,1). Ce dernier s'est notamment révélé meilleur que ceux qui sélectionnent la complexité du modèle avec un critère comme l'AIC.

Ce meilleur modèle a été appliqué pour prédire l'incidence du cancer en Suisse jusqu'en 2025, anticipant une stabilisation du risque de développer un cancer dans les années à venir. En combinant cette tendance avec les projections démographiques de l'Office fédéral de la statistique, nous avons cependant anticipé une augmentation substantielle du nombre annuel de nouveaux cas de cancer, entièrement due aux changements démographiques. Cette augmentation a été estimée à +18% pour les hommes et +11% pour les femmes, avec des augmentations allant de 4,15% pour la thyroïde chez l'homme à 26% pour la vessie chez l'homme.

L'estimation (et la prévision) des tendances de l'incidence du cancer au fil du temps peut être en partie faussée par les changements dans les processus de détection du cancer, tels que, mais sans s'y limiter : l'introduction ou la modification des programmes de dépistage, l'utilisation de différents outils de dépistage, et la détection opportuniste. Dans la troisième partie de cette thèse, nous avons proposé un modèle capable de s'ajuster à ces changements et donc d'estimer la véritable tendance sous-jacente de l'incidence du cancer.

Introduction

The main theme of this PhD thesis was to predict the evolution of cancer incidence in Switzerland. This was part of an interdisciplinary project supported by the Swiss Cancer League. It combined statistical, epidemiological and public health aspects, and used cancer registry data from Vaud, Geneva, Neuchâtel, and all Swiss registries combined. The thesis is divided into three parts, each of which has resulted in a scientific article, the first two being already published, and the third almost ready for submission. The three articles are attached to this thesis. It should be noted that the article related to the second part was published in 2022, before the article related to the first part, which was published in 2023.

The starting point of this research was an application of a commonly used method for making predictions of the age-standardized incidence of prostate cancer in Switzerland after 2010, based on data observed between 1982 and 2010. The result was surprising (Figure 1). As can be seen in this figure, there was a trend reversal around the year 2005, with the incidence increasing before that year, and decreasing thereafter, as shown by the blue dots in this graphic. Yet, one of the most popular methods to model cancer incidence (an APC model extrapolating a drift only, see the first article for more explanations) predicted an abrupt and certainly unrealistic increase of the incidence after 2010, as shown by the solid cyan line. Another method (an APC model extrapolating all effects) produced very different predictions, continuing the downward trend (see the dashed green line). In fact, it turned out that such a scenario with a trend reversal was the most difficult one to make accurate predictions. In Figure 1, each grey line shows a prediction obtained with a different method, all implemented in the literature on cancer incidence. It can be seen that these predictions were extremely different from each other, raising the crucial and inevitable question of which method to use.

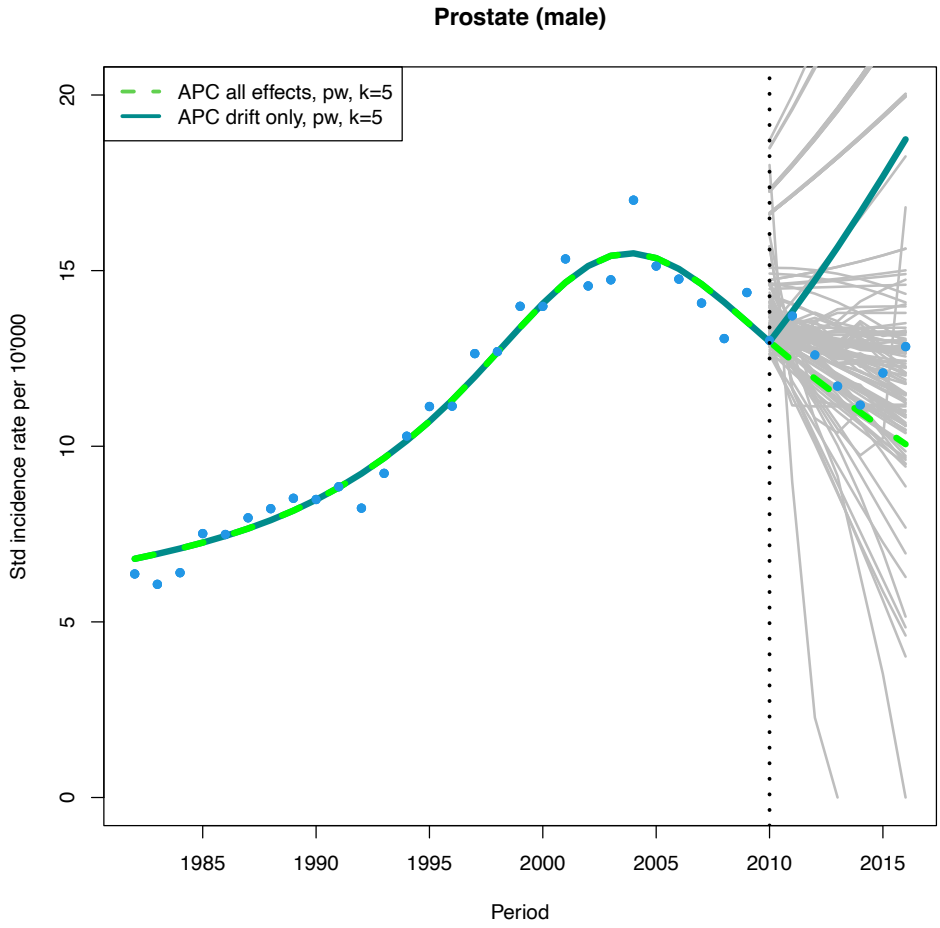

Figure 1 : Prostate age-standardized cancer incidence in blue points, models compared are estimated with data from 1982 to 2010, predictions are produced with all the methods from 2011 to 2016; highlighted lines in green and dark cyan are two extrapolation variants of the most used model.

First part

The first part of this thesis, published in Trächsel et al. (2023), was to review and compare performances of models and methods used in the statistical and epidemiological literature to predict cancer incidence.

Numerous prediction methods have been employed in the field of cancer incidence, including age-period-cohort (APC) models (Carstensen, 2007; Holford, 1983; Rutherford et al., 2012), Bayesian age-period-cohort (BAPC) models (Riebler et al., 2012; Riebler & Held, 2017; Schmid & Held, 2004), autoregressive integrated moving average (ARIMA) time-series models (Hamilton, 1994), neural networks (Hastie et al., 2009), joinpoint regression (Lerman, 1980), Poisson and negative binomial generalized linear models (GLM) (Cameron & Trivedi, 1998), state space models (Chen et al., 2012;

Hamilton, 1994), and vector autoregressive Hilbert–Huang transform (Huang & Attoh-Okine, 2005), among others. A recent systematic review of lung cancer incidence prediction studies (Yu et al., 2019) indicated that APC, BAPC, GLM, and joinpoint were the most frequently used methods. Most of these methods can also be applied using different variants, including various possibilities for the priors, tuning parameters, extrapolation methods or model complexity, making the choice even richer.

Although there have been some comparisons between prediction models in the literature, these were generally limited to a few models, or different variants of a same class of models (Chen et al., 2012; Clements, 2005; B. Møller et al., 2003; Riebler et al., 2012; Zheng et al., 2020). Most of these comparisons were also conducted in situations where incidences were almost constantly increasing, representing a relatively unchallenging prediction setting. In recent years in Switzerland, however, the incidence of certain cancers (e.g., breast, prostate, and skin melanoma;(OFS, 2020)) have started stabilizing or decreasing, as in Figure 1. This adds complexity to the task of predicting incidence trends and highlights the importance of comparing the different methods in this new challenging context.

In this first part of the thesis, we thus compared the performance of 165 (variants of) models to predict cancer incidence, the most important comparative study to date. We included 70 GLM (including joinpoint regression), 12 APC (including different extrapolation methods), 10 BAPC (including different priors), 65 ARIMA (with different orders of complexity) and 8 linear models (LM), where we simply extrapolated regression lines. All these methods were compared over 150 scenarios obtained from real cancer incidence data at different sites (e.g., breast, prostate, colorectal, lung, skin melanoma) from the registries of Vaud, Geneva and Neuchâtel, representing many kinds of realistic situations, including trend reversals. Models were estimated (or "trained") on one part of the data to make predictions on the other part, and they could be compared with respect to the "truth" which was thus known.

Implementing all methods proved to be challenging. In particular, Bayesian methods were extremely time-consuming, with a total computation time of 33 days on a machine with 8 cores to make predictions for the 150 scenarios.

Second part

In the second part of this thesis, published in Trächsel et al.(2022), we applied the best method identified in the first part (a low-order ARIMA model) to make predictions of cancer incidences in Switzerland up to 2025.

Predictions were made for the age-standardized incidences of the 12 most frequent cancer sites according to the ICD10 codes (Fritz et al., 2000): oral cavity and pharynx (C00-14), stomach (C16), colon and rectum (C18-20), lung-bronchus-trachea (C33-34), skin melanoma (C43), breast (C50), corpus uteri and uterus NOS (C54-55), ovary (C56), prostate (C61), bladder (C67), thyroid (C73), non-Hodgkin lymphoma (C82-85, C96), separately for women and men, when applicable. We also made a prediction for the other cancers in an “other cancers” category to be able to predict the total number of cancers. As done in the literature (Møller et al., 2007), also for Switzerland (Rapiti et al., 2014), predictions of age-standardized cancer incidence rates were then combined with the population projections of the Swiss Federal Statistical Office (baseline scenario "Average 10-2020") to obtain a prediction of the future number of new cancer cases.

It is usual in the literature (H. Møller et al., 2007; Rapiti et al., 2014) to decompose a change (e.g. an increase) in the number of new cancer cases over time in two components, one due to the change in the epidemiological risk of getting cancer, and another one due to the demographic changes in the population. One originality of our work was to decompose such a change not in two, but in three components: an epidemiological component, via the evolution of the age-standardized cancer incidence rates, capturing the possible changing exposure at the population level to the different risks of getting cancer, and two demographic components, namely population growth and population aging. This decomposition technique enabled us to emphasize the role of these three components, which might be useful for planning, or to evaluate the potential impact of prevention campaigns on cancer incidence, which will typically affect only the first component.

Third part

The third part of this work is the most original one, with the ambitious and difficult goal of estimating cancer incidence while taking into account (i.e. adjusting for) the changes in the cancer detection processes such as but not limited to: the introduction or modification of screening programs, the use of different screening tools, and incidental detection. An article is currently in the process of being prepared for submission.

A few studies have explored the relationship between screening and cancer incidence, seeking to distinguish trends that are due to a real change in cancer incidence in the population from those that are due to changes in screening practice, such as an increase in the frequency of screening which would inevitably lead to the detection of more new cases of cancer. This might be particularly relevant for prostate cancer detected by the prostate-specific antigen (PSA) screening test. For instance, Telesca (2008) developed a likelihood model, employing PSA utilization over time to simultaneously estimate the mean lead time (the period between diagnosis via screening and the time cancer would have been clinically diagnosed with symptoms) and a smooth secular trend (the one that would have occurred without the introduction of PSA tests). In another study, Etzioni (2008) investigated the impact of PSA screening on prostate cancer incidence separately for low- and high-stage cancers, utilizing multiple models and assumptions, including a simulation model of the natural progression of cancer through pathologic stages until clinical (symptomatic) detection (Cowen, 1994), and a second simulation model of PSA growth rate. Gulati (2014) employed similar models to predict past and future cancer overdiagnosis (cancers that would not have been detected in the absence of screening) under continued or discontinued PSA screening programs.

The link between cancer screening and cancer stage distribution (from stage I to stage IV) has also been examined in the literature. For example, Toyoda (2020) investigated the stage distribution and the proportion of screen-detected cases for various cancers using Osaka Cancer Registry data (Japan), discovering a strong association between high proportions of screen-detected cases and high proportions of low-stage cancers. Cardoso (2022) reached a similar conclusion in Europe based on data from 16

population-based cancer registries, revealing significantly higher proportions of low-stage cancers (stages I and II) among screen-detected cases compared to non-screen-detected cases.

Rather than utilizing a simulation model of cancer progression, or precise data on changes in screening practice over time which would be difficult to obtain, our idea was to postulate and to exploit that, in the absence of change in the cancer detection, the distribution of cancer stages would remain constant (while the level of incidence may, of course, vary). Consequently, we interpret changes in this distribution over time as a proxy for changes in detection. Although challenging to test, this assumption is consistent with the results of Toyoda (2020) and Cardoso (2022), which demonstrate a strong association between screening detection and cancer stage. Loosely using the term “screening” to refer to such changes, we thus proposed a “Screening Adjusted Model” (SA model), which includes the real trend to be estimated, and the mean of the stage distribution as a confounding variable, lagged to a certain time into the past. In order to reduce the number of coefficients in the model, and to get credible estimates, we also imposed some structure and constraints on the coefficients. We evaluated the performance of the SA model, in comparison with a classical model without adjustment, via simulated data according to various scenarios of changes in cancer detection that could be produced by changes in the screening practice over time and different real secular trends, where the goal was to get an estimate of these trends. We also applied our model to breast cancer data from the registry of Geneva.

In the subsequent sections, we briefly present the results obtained in the three parts of the thesis, before discussing their implications, mentioning some potential future work and presenting our concluding thoughts. The three articles constitute the final portion of this manuscript. Supplementary materials for these articles are provided in a separate document.

Results

In this section, we briefly present the results from the three parts of this thesis described in the Introduction, that are related to the three articles attached at the end of this document.

First part

In the first article, we compared the performance of 165 (variants) of models to predict age-standardized cancer incidence. We found that the simplest models, such as ARIMA of order (2,1,1) or (1,1,0) and linear models (LM) that simply extrapolate a trend estimated over the last few years, were the most effective methods for predicting the evolution of age-standardized incidence rates, both in terms of point predictions and in terms of the quality (width and coverage rates) of the prediction intervals. More complicated and widely used models, such as APC and BAPC, did not perform as well, while GLMs were found to be the worse models for prediction. Methods based on the well-known AIC criterion to select a model also showed poor performances.

Unsurprisingly, the most difficult scenarios for prediction were some with a reversal trend, such as the situation presented in Figure 1. This is where we observed the greatest disparity between the predictions produced by the different methods, as illustrated in this figure. The point is that simple models, such as LMs, were able to quickly adapt to new trends, while others, like APC, struggled to do so, especially the APC model extrapolating only the drift, which is the most widely adopted to date.

Second part

In the second article, we predicted the number of new cases of cancer in Switzerland for the years 2019-2025, for the 12 most frequent cancer sites, separately for women and men. Predictions of age-standardized incidence was done based on data from years 1987-2018 using an ARIMA of order (2,1,1), the best method according to our first article above, which were then combined with demographic projections from the Swiss Federal Statistical Office.

According to our model, an overall stabilization of age-standardized incidence is expected in Switzerland through 2025 for both sexes, with a projected decrease of less than 1%. For Swiss men, slight increases of incidences are anticipated for bladder cancer and skin melanoma, while slight decreases are expected for non-Hodgkin lymphoma, lung, bronchus, and trachea, colon-rectum, and stomach, with more significant declines for oral cavity & pharynx and thyroid. Prostate cancer incidence is expected to remain relatively constant. For Swiss women, slight increases of incidences are predicted for cancers of the oral cavity & pharynx and thyroid, while minor decreases are expected for cancers of the colon-rectum, breast, and uterus, with more substantial declines for ovarian, lung, bronchus, trachea, bladder, and stomach cancers. No changes in incidence are anticipated for cutaneous melanoma and non-Hodgkin's lymphoma.

Although age-standardized cancer incidence is expected to remain relatively stable up to 2025 for most sites, the number of new cancer cases will increase substantially due to demographic trends. The total number of new cancer cases is expected to rise by 18% for men and 11% for women, resulting in a 15% overall increase between 2018 and 2025. This growth is entirely attributable to the aging and growth of the Swiss population, i.e., to the demographic components, not to the epidemiological component. Similarly, predicted increases in the number of new cancer cases for any site are primarily due to the demography. The ranking of the three most frequent cancers for men (prostate, lung, colon-rectum) will remain unchanged in 2025, while colorectal cancer is expected to become the second most frequent cancer for women in 2025, after breast cancer, overtaking lung cancer.

Third part

In the third article, we propose a Screening Adjusted (SA) model to estimate a trend in incidence rate adjusted for (i.e., that would have been observed without) changes in cancer detection, using the mean stage distribution as a proxy. Our simulations showed that the SA model provided estimates of incidence rates that were closer to those that would be observed in the absence of any changes in the detection than when using a classical (e.g., APC) model. The SA model is thus particularly useful for analyzing trends of cancer data when, for example, screening practices vary over time, without the need of knowing

details about when and how these changes occurred. An application of the SA model to estimate incidence trends of breast cancer in the canton of Geneva illustrated how the SA estimate may differ from the classical one.

Discussion

The main theme of this thesis is the comparison of statistical models for the projection of future cancer incidence, an extremely important task in modern societies where cancer has unfortunately become a major public health problem and cause of death. The choice of one projection method can produce very different results and is therefore crucial for obtaining sound and useful predictions for policy-makers. By exploiting the data from the registers of the cantons of Vaud, Geneva, and Neuchâtel (the oldest registers in Switzerland), and using the cross-validation method on these data, we were able to compare the predictive performance of a very large number of prediction methods and models (variants), both in terms of point prediction accuracy and quality of the prediction intervals. Our findings contrast somewhat with the literature on the subject (Chen et al., 2012; Clements, 2005; B. Møller et al., 2003; Riebler et al., 2012; Zheng et al., 2020). While in general complex or even very complex methods are used, based on the estimation of latent effects such as age, period and/or cohort effects, or attempting to select the complexity of the model using a criteria such as the AIC (Akaike Information Criterion), we have shown the superiority of using simpler methods such as ARIMA models or classical linear models, which simply extrapolate and smooth the most recent observed trends. These methods, less used until now in cancer incidence prediction, have indeed proved to be more suitable to adapt to the recent stabilizations and trend reversals observed in the Swiss incidence data.

The best model from this large comparative study, a low-order ARIMA model, was then used to predict the future evolution of cancer incidence in Switzerland. This model was then combined with the population projections of the Swiss Federal Statistical Office (FSO) to obtain a prediction of the evolution of the annual number of new cancer cases, which is information more directly exploitable by the decision maker. Our results indicate a stabilization of the age-standardized incidence of most cancers in the coming years, accompanied by a significant increase in the number of new cases each year, due

to the increase and aging of the population according to the FSO projections. Indeed, we expect an 18% increase in new cancer cases among men and 11% among women, with a particularly strong increase in colorectal cancer among the latter, which would take second place after breast cancer in 2025, replacing lung cancer.

Since the estimation and extrapolation of cancer incidence trends over time can be affected by changes in detection processes (screening programs, detection techniques, opportunistic detection, etc.), in the last part (part 3) of the thesis we also worked on the development of a new model able to correct the observed incidence for these changes and thus able to estimate the "true" pattern of cancer incidence (at constant detection). As all factors affecting the detection process are very difficult to measure, we used as a proxy the distribution of detected cancer stages, from stage I to stage IV. Based on several studies that have demonstrated the relationship between screening and cancer stage (Cardoso et al., 2022; Toyoda et al., 2020), we assumed that this distribution remains roughly stable if no change in the detection process occurs, and that it will change, e.g. showing a lower mean stage, if a change in the detection process occurs, e.g. the introduction of a new screening practice. Our method therefore consists in introducing into a classical model of incidence over time the effect of year-to-year variations in the mean stage. These variations are lagged several times in the past, since a change in the mean stage in a given year produces effects over several years, which gradually disappear over time. Our model was tested on a simulation study mimicking different scenarios with more abrupt or more gradual changes in the detection process (and thus in the mean stage) and with different underlying incidences. This is the most innovative part of this thesis, an idea that is still not perfect, but which seems to us to be promising, although with room for improvement.

The main implication of our study considered as a whole is therefore a recommendation, for those interested in prediction in the field of cancer, to avoid using complex methods for this purpose (those most commonly adopted at present). These models, such as the Age-Period-Cohort (APC) models or their Bayesian counterparts (BAPC models), which produce estimates that are smooth and close to the observed data, struggle to adapt to the stabilization and trend changes observed for most cancers in

recent years. These models are therefore useful for interpreting observed trends but should be avoided for projections. The best models according to our comparative study are those that simply extrapolate the latest trends, such as the classical linear models, with the addition of a smoothing of these trends, such as the ARIMA models. In case of stabilization of the incidence curves, reversal of the trends or simply erratic behavior in the last observation period, the ARIMA methods will return a "conservative" prediction in the sense of future stabilization, as we obtained for Switzerland. This latter prediction turns out to be better than that of the (B)APC models, which is sometimes aberrant as shown in Figure 1. The ARIMA models were indeed better not only because they produced predictions closer to the true values than the other methods (point accuracy), but also in terms of coverage of the prediction intervals, the closest to the nominal value of 95%. Having credible prediction intervals around the future incidence is a great advantage and a very important piece of information for those who have to make decisions in terms of resource allocation.

Our results are certainly related to the data used to test the different models, namely incidence data in three Swiss cantons. However, these data reflect quite well a trend observed more generally in modern societies in recent years, with (fortunately) incidences stopping increasing and often entering periods of stabilization or changing trend. It is easy to imagine that if the study had been conducted in a context of steadily increasing cancer incidence (as was the case a few decades ago for most cancers), the results in terms of model rankings might have been different, but in that case all the models would have predicted about the same thing (a continuation of that increase) and the choice of the "right" model would have been less crucial.

One can, of course, imagine extending our comparative study to other models or model variants. For example, in the context of GLM models, we have limited ourselves to the Poisson distribution and the logarithmic link function. In the framework of Bayesian APC models, we have considered five different priors, the ones most used in the literature. An overdispersion could be introduced in the Poisson model and the latter could eventually be replaced by the Negative Binomial distribution in the GLM, as other priors could be tested in the framework of Bayesian models. The number of possibilities is of course

infinite. However, having tested and compared a total of 165 model variants, we do not believe that such variations would be able to change the main message of our study.

Another downside of the prediction models, both the ARIMA models and the more complex models estimating age, period and cohort effects, is that they do not include additional information about the evolution of the exposure to carcinogens and behavioral and environmental risk factors. It would indeed be very interesting to take into account this kind of information, but the disadvantage in the context of a projection is that it would be necessary to combine the projection of incidence with that of these same factors, by introducing an additional element of uncertainty into the projections. A previous attempt to introduce additional information on tobacco consumption into the prediction of lung cancer incidence has been shown to be unable to improve the prediction (Knorr-Held, 2001).

A further very important point for the decision maker is the possibility of getting predictions not only of the future age-standardized incidence of the different cancers, but also of the absolute number of new cases. To project new cases, as mentioned, one must combine the expected evolution of incidence with population projections, as we have done for Switzerland in our second publication and as is also done in general in the literature (H. Møller et al., 2007; Rapiti et al., 2014). Having used an ARIMA model to project age-standardized cancer incidence, and differently from the predictions made by an APC model, we do not dispose of age-specific predictions of incidence. For this reason, combining incidence with population projections required the assumption that the distribution of cancers across age groups remains the same over time (which was recognized as a limitation in our paper). This assumption made it possible to decompose the expected future evolution of cancer cases into three components, the first epidemiological (incidence) and the other two demographic, population increase and aging. Aging was found to be more important than increasing population size as a driver of the future number of new cancer cases, particularly among men.

Another important limitation in this context, a limitation shared so far by all studies that have combined incidence with population projections to obtain a projection of new cases (at least to our knowledge), is

the impossibility of obtaining prediction intervals, because the population projections carried out by the FSO (and in general by the Statistical Offices) do not currently produce such intervals for the expected evolution of the population. Instead, scenarios of evolution are provided, and we consider the average scenario, which does not include a measure of uncertainty. This aspect therefore deserves to be studied further, with the development of population projection techniques that include an estimate of uncertainty. The uncertainty in the age-standardized incidence (which we have calculated) might then be combined with the estimate of the uncertainty in future demographic trends (which is not yet available) by suitable statistical techniques, such as bootstrapping or Monte Carlo simulations.

Finally, the possibility of obtaining an estimate (and eventually a prediction) of incidence adjusted for changes in the detection process over time was surely our most ambitious goal. Although several articles exist in the literature in this field, to our knowledge, we are the first to have exploited the link between changes in this process (e.g. a new prevention campaign) and the distribution of detected cancer stages (e.g. a lower mean stage), in order to make this correction with the minimum of information. Thus, we only need to know how the mean stage changes over time to obtain an adjusted estimate of incidence. However, this information is not as easy to obtain as one might imagine. Indeed, at the time of registration of a new cancer, information on the stage of the cancer (I-IV) has only been collected by some registries and for some cancer sites (notably breast cancer). Moreover, if these registries (Vaud and Geneva) have already been active in 1982, the registration of the stage only becomes reliable for some sites from 1991 onwards, when the number of missing data stabilizes at a percentage lower than 10%. The applicability of our model, which we were able to test extensively on simulated data, is therefore considerably limited to certain cantons, certain sites and certain years, by the availability of the data. In addition, as our model is based on lagged data (due to the long-lasting effect on the incidence of a change in the mean stage), the number of years over which an estimate is possible is further reduced. This aspect represents a limitation of our (third) study. In particular, the lack of longer data series prevented us from conducting a sensitivity analysis comparing different choices of lagged periods and forced us to summarize the distribution of stages with its mean in order to avoid too many parameters compared to the data available. Possible solutions could be sought in the aggregation of several

international sources (but still the percentage of missing values must be comparable in the different sources) to improve the estimation of the mean stage effect or the analysis of semi-annual data to increase the length of the series available.

Another important limitation of our model is the implicit assumption that changes in the detection process do not alter the underlying incidence. While this is true for most cancers, since, for instance, an increase in screening will only anticipate a detection that would have occurred later, this is not the case for those cancers for which an early detection of precancerous lesions (and their subsequent treatment) prevents the occurrence of cancer and therefore reduces the incidence. This is notably the case for colorectal cancer, to which our model does not apply in its current form. Our model could be adapted, by considering the precancerous lesions as part of the incidence.

In conclusion, modeling and forecasting cancer incidence presents significant challenges. Given the limited availability of useful covariates, our capabilities are often limited to extrapolating the most recent observed trends. Our study highlighted the predominant role of demographic changes in determining future cancer incidence, population aging and growth appearing to be the main drivers in this regard. Finally, we have developed a method to adjust past incidence trends for changes in the detection process. We hope that this method will allow us to improve prediction, for example by applying an ARIMA model not to the observed incidence but to the incidence that has been adjusted for changes in the detection process. This approach has great potential for expansion and refinement. I am personally very interested in contributing to its future development, such as the development of an R package and the application of this model to multiple series of cancer incidence to facilitate its adoption by the scientific community.

References

Cameron, & Trivedi. (1998). *Regression Analysis of Count Data*. Cambridge University Press. https://doi.org/10.1017/CBO9780511814365

Cardoso, R., Guo, F., Heisser, T., De Schutter, H., Van Damme, N., Nilbert, M. C., Christensen,

J., Bouvier, A.-M., Bouvier, V., Launoy, G., Woronoff, A.-S., Cariou, M., Robaszkiewicz, M., Delafosse, P., Poncet, F., Walsh, P. M., Senore, C., Rosso, S., Lemmens, V. E. P. P., … Brenner, H. (2022). Overall and stage-specific survival of patients with screen-detected colorectal cancer in European countries: A population-based study in 9 countries. *The Lancet Regional Health - Europe*, *21*, 100458. https://doi.org/10.1016/j.lanepe.2022.100458

Carstensen, B. (2007). Age–period–cohort models for the Lexis diagram. *Statistics in Medicine*, *26*(15), 3018–3045. https://doi.org/10.1002/sim.2764

Chen, Portier, K., Ghosh, K., Naishadham, D., Kim, H.-J., Zhu, L., Pickle, L. W., Krapcho, M., Scoppa, S., Jemal, A., & Feuer, E. J. (2012). Predicting US and state-level cancer counts for the current calendar year: Part I: evaluation of temporal projection methods for mortality. *Cancer*, *118*(4), 1091–1099. https://doi.org/10.1002/cncr.27404

Clements, M. S. (2005). Lung cancer rate predictions using generalized additive models. *Biostatistics*, *6*(4), 576–589. https://doi.org/10.1093/biostatistics/kxi028

Cowen, M. (1994). A Markov model of the natural history of prostate cancer. *Journal of Clinical Epidemiology*, *47*(1), 3–21. https://doi.org/10.1016/0895-4356(94)90029-9

Etzioni, R., Gulati, R., Falcon, S., & Penson, D. F. (2008). Impact of PSA Screening on the Incidence of Advanced Stage Prostate Cancer in the United States: A Surveillance Modeling Approach. *Medical Decision Making*, *28*(3), 323–331. https://doi.org/10.1177/0272989X07312719

Fritz, A., Percy, C., Jack, A., Shanmugaratnam, K., Sobin, L. H., Parkin, D. M., Whelan, S. L., & World Health Organization. (2000). International classification of diseases for oncology. *Classification internationale des maladies pour l'oncologie*. WHO IRIS. https://apps.who.int/iris/handle/10665/42344

Gulati, R., Tsodikov, A., Etzioni, R., Hunter-Merrill, R. A., Gore, J. L., Mariotto, A. B., & Cooperberg, M. R. (2014). Expected population impacts of discontinued prostate-specific antigen screening. *Cancer*, *120*(22), 3519–3526. https://doi.org/10.1002/cncr.28932

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691042893/time-series-analysis

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Holford, T. R. (1983). The Estimation of Age, Period and Cohort Effects for Vital Rates. *Biometrics*, *39*(2), 311–324. JSTOR. https://doi.org/10.2307/2531004

Huang, N. E., & Attoh-Okine, N. O. (Eds.). (2005). *The Hilbert-Huang transform in engineering*. Taylor & Francis.

Knorr-Held, L. (2001). Projections of lung cancer mortality in West Germany: A case study in Bayesian prediction. *Biostatistics*, *2*(1), 109–129. https://doi.org/10.1093/biostatistics/2.1.109

Lerman, P. M. (1980). Fitting Segmented Regression Models by Grid Search. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *29*(1), 77–84. JSTOR. https://doi.org/10.2307/2346413

Møller, B., Fekjaer, H., Hakulinen, T., Sigvaldason, H., Storm, H. H., Talbäck, M., & Haldorsen, T. (2003). Prediction of cancer incidence in the Nordic countries: Empirical comparison of different approaches: Comparison of Methods for Incidence Prediction. *Statistics in Medicine*, *22*(17), 2751–2766. https://doi.org/10.1002/sim.1481

Møller, H., Fairley, L., Coupland, V., Okello, C., Green, M., Forman, D., Møller, B., & Bray, F. (2007). The future burden of cancer in England: Incidence and numbers of new patients in 2020. *British Journal of Cancer*, *96*(9), 1484–1488. https://doi.org/10.1038/sj.bjc.6603746

OFS. (2020, November 12). *Cancer, nouveaux cas et décès: Nombre, taux et évolution par localisation cancéreuse et période - 1987-2017 | Tableau*. Office fédéral de la statistique. https://www.bfs.admin.ch/asset/fr/14816247

Rapiti, E., Guarnori, S., Pastoors, B., Miralbell, R., & Usel, M. (2014). Planning for the future: Cancer incidence projections in Switzerland up to 2019. *BMC Public Health*, *14*(1), 102. https://doi.org/10.1186/1471-2458-14-102

Riebler, A., & Held, L. (2017). Projecting the future burden of cancer: Bayesian age-period-cohort analysis with integrated nested Laplace approximations: Projecting the future burden of cancer. *Biometrical Journal*, *59*(3), 531–549. https://doi.org/10.1002/bimj.201500263

Riebler, A., Held, L., & Rue, H. (2012). Estimation and extrapolation of time trends in registry data—Borrowing strength from related populations. *The Annals of Applied Statistics*, *6*(1). https://doi.org/10.1214/11-AOAS498

Rutherford, M. J., Thompson, J. R., & Lambert, P. C. (2012). Projecting Cancer Incidence using Age-period-cohort Models Incorporating Restricted Cubic Splines. *The International Journal of Biostatistics*, *8*(1). https://doi.org/10.1515/1557-4679.1411

Schmid, V., & Held, L. (2004). Bayesian Extrapolation of Space-Time Trends in Cancer Registry Data. *Biometrics*, *60*(4), 1034–1042. JSTOR.

Telesca, D., Etzioni, R., & Gulati, R. (2008). Estimating Lead Time and Overdiagnosis Associated with PSA Screening from Prostate Cancer Incidence Trends. *Biometrics*, *64*(1), 10–19. https://doi.org/10.1111/j.1541-0420.2007.00825.x

Toyoda, Y., Tabuchi, T., Hama, H., Morishima, T., & Miyashiro, I. (2020). Trends in clinical stage distribution and screening detection of cancer in Osaka, Japan: Stomach, colorectum, lung, breast and cervix. *PLOS ONE*, *15*(12), e0244644. https://doi.org/10.1371/journal.pone.0244644

Trächsel, B., Rapiti, E., Feller, A., Rousson, V., Locatelli, I., & Bulliard, J.-L. (2022). Predicting the burden of cancer in Switzerland up to 2025. *PLOS Global Public Health*, *2*(10), e0001112. https://doi.org/10.1371/journal.pgph.0001112

Trächsel, B., Rousson, V., Bulliard, J.-L., & Locatelli, I. (2023). Comparison of statistical models to predict age-standardized cancer incidence in Switzerland. *Biometrical Journal*, 2200046. https://doi.org/10.1002/bimj.202200046

Yu, X. Q., Luo, Q., Hughes, S., Wade, S., Caruana, M., Canfell, K., & O'Connell, D. L. (2019). Statistical projection methods for lung cancer incidence and mortality: A systematic review. *BMJ Open*, *9*(8), e028497. https://doi.org/10.1136/bmjopen-2018-028497

Zheng, Y., Zhang, L., Zhu, X., & Guo, G. (2020). A comparative study of two methods to predict the incidence of hepatitis B in Guangxi, China. *PLOS ONE*, *15*(6), e0234660. https://doi.org/10.1371/journal.pone.0234660

Received: 16 February 2022 | Revised: 7 February 2023 | Accepted: 1 March 2023

DOI: 10.1002/bimj.202200046

Biometrical Journal

RESEARCH ARTICLE

Comparison of statistical models to predict age-standardized cancer incidence in Switzerland

Bastien Trächsel | **Valentin Rousson** | **Jean-Luc Bulliard** | **Isabella Locatelli**

Center for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne, Switzerland

Correspondence
Bastien Trächsel, Center for Primary Care and Public Health (Unisanté), University of Lausanne, Route de Berne 113, 1010, Lausanne, Switzerland.
Email: sui.bastien@gmail.com

Funding information
Krebsliga Schweiz, Grant/Award Number: KFS-4385-02-2018

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article were reproduced partially due to data confidentiality issues.

Abstract

This study compares the performance of statistical methods for predicting age-standardized cancer incidence, including Poisson generalized linear models, age-period-cohort (APC) and Bayesian age-period-cohort (BAPC) models, autoregressive integrated moving average (ARIMA) time series, and simple linear models. The methods are evaluated via leave-future-out cross-validation, and performance is assessed using the normalized root mean square error, interval score, and coverage of prediction intervals. Methods were applied to cancer incidence from the three Swiss cancer registries of Geneva, Neuchatel, and Vaud combined, considering the five most frequent cancer sites: breast, colorectal, lung, prostate, and skin melanoma and bringing all other sites together in a final group. Best overall performance was achieved by ARIMA models, followed by linear regression models. Prediction methods based on model selection using the Akaike information criterion resulted in overfitting. The widely used APC and BAPC models were found to be suboptimal for prediction, particularly in the case of a trend reversal in incidence, as it was observed for prostate cancer. In general, we do not recommend predicting cancer incidence for periods far into the future but rather updating predictions regularly.

KEYWORDS

age-period-cohort models, age-standardized cancer incidence, autoregressive integrated moving average, Bayesian age-period-cohort models, generalized linear models, interval score, prediction interval, root mean square error, trend reversal

1 | INTRODUCTION

Cancer has progressively become a major public health problem in Western countries like Switzerland where, each year, about 40,000 new cases are diagnosed and 17,000 induced deaths occur (OFS, 2020). Cancer is the second leading cause of mortality in Switzerland, and the main cause in females aged 25–84 years, and in males aged 45–84 years (OFS, 2020). Prediction of cancer incidence is an important public health issue in modern societies to plan resource allocation and prevention interventions and to help address future cancer research.

Many prediction methods are used in the field of cancer incidence, such as age-period-cohort (APC) (Carstensen, 2007; Holford, 1983; Rutherford et al., 2012), and Bayesian age-period-cohort (BAPC) (Riebler et al., 2012; Riebler & Held, 2017;

Schmid & Held, 2004) models, autoregressive integrated moving average (ARIMA) time series models (Hamilton, 1994), neural networks (Hastie et al., 2009), joinpoint regression (Lerman, 1980), Poisson and negative binomial generalized linear models (GLM) (A. C. Cameron & Trivedi, 2013a), state space models (H. S. Chen et al., 2012; Hamilton, 1994), and the vector autoregressive Hilbert–Huang transform (H. S. Chen et al., 2012). In a recent systematic review of 101 published studies on lung cancer incidence prediction, the most commonly used methods were APC, BAPC, GLM, and joinpoint (Yu et al., 2019). APCs were largely applied to predict the incidence of common cancer sites (Coupland et al., 2010; Møller et al., 2002; Rapiti et al., 2014) as well as BAPC (J. K. Cameron & Baade, 2021; W.-Q. Chen et al., 2011; Shi et al., 2021), GLM (NSW Cancer Institute, 2016; Yang et al., 2005; Zemni et al., 2022), and joinpoint (Lin et al., 2021; Rahib et al., 2021; Wong et al., 2021). Other studies used ARIMA models (Earnest et al., 2019; Li et al., 2022; Tsoi et al., 2017).

Comparisons among prediction models have been carried out in the literature. For example, Møller (2003) compared several variants of APC predicting the incidence of the most frequent cancer sites and concluded that APC models using a power link function were superior to those adopting the canonical logarithmic link. To predict the incidence of lung, bronchus, and trachea cancers, Riebler (2012) compared BAPC with the Lee–Carter model, a model used in the demographic field to predict mortality (Lee & Carter, 1992), showing the superiority of BAPC. ARIMA has been compared to neural networks showing the similar performance (Zheng et al., 2020). Using USA state-level data to predict the incidence of the most frequent cancer sites, Chen (2012) compared the performance of APC, BAPC, joinpoint regression, state space models, and the vector autoregressive Hilbert–Huang transform and concluded the superiority of BAPC, although followed closely by APC and state space models. In another study (Clements, 2005), generalized additive models (GAM, a special case of GLM) have shown a better performance than BAPC to predict the incidence of lung cancer.

These comparisons were in general limited to a few models only (e.g., BAPC vs. Lee-Carter, Riebler et al., 2012; or BAPC vs. GAM, Clements, 2005) or to different variants of the same class of models, for example, APC (Møller et al., 2003), with the notable exception of the U.S. study (H. S. Chen et al., 2012), which, however, omitted important methods such as GLM or ARIMA. In addition, most of these comparisons have been made in situations where the incidence was almost constantly increasing, providing a relatively easy setting for prediction. In recent years in Switzerland, we observed some trend reversals, as the incidence of some cancers (e.g., breast, prostate, and skin melanoma) has begun to stabilize or decrease (OFS, 2020). This phenomenon represents an additional difficulty in predicting incidence trends, and it is of interest to compare the different methods in this context.

In the present paper, we sought to compare a large number of methods for predicting the incidence of the most common cancers, including in situations with a trend reversal, based on cancer incidence data from Switzerland. The methods compared include Poisson GLM, of which the Lee–Carter model and joinpoint regression are special cases, APC and BAPC models, ARIMA, and simple linear regression models. We evaluated prediction performance by repeatedly using leave-future-out cross-validation, that is, dividing the data into a training set (on which the models are fitted) and a test set (on which the predictions are evaluated) and considering all possible period partitions of the data in these two sets. This allowed us to produce many prediction settings (scenarios), including some with a trend reversal at the boundary of the training set.

This paper is organized as follows. The data used for our comparisons are presented in Section 2. Section 3 explains how we evaluated the performance of a prediction method by leave-future-out cross-validation using different criteria. All compared methods are described in Section 4. Our results are detailed in Section 5, and a discussion is proposed in Section 6.

2 | DATA DESCRIPTION

2.1 | Data sources

Our primary data sources were population-based Swiss cancer registries of the cantons of Vaud, Geneva, and Neuchatel. Registries record information on all incident cases of malignant neoplasms occurring in their resident population according to international rules, such as the International Classification of Diseases for Oncology (Fritz et al., 2000). For each incident case, information on the incidence date, the patient's date of birth, and the cancer site are recorded, among others.

We used combined data from the three cancer registries of the cantons of Vaud, Geneva, and Neuchatel because they are the oldest registries in Switzerland. Incidence records were available from 1982 to 2016, the latter year being the most recent with complete data. We considered the five most frequent cancer sites: breast, colorectal, lung, prostate, and skin melanoma and formed a final group including all other sites. Considering separately the two sexes, we thus dealt with

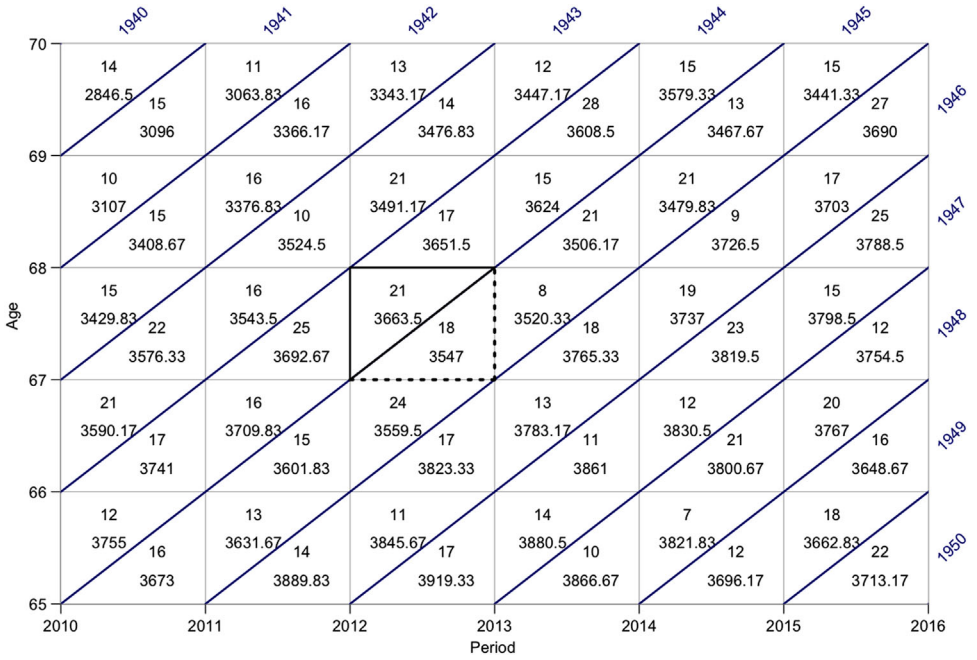

FIGURE 1 Lexis diagram of an extract of the combined data from the Swiss registries of Vaud, Geneva, and Neuchatel. Selected period: January 1, 2010–December 31, 2015. Selected age range: 65th to 70th birthday. In each triangle, the top number represents the number of incident cases $D_{a,p,c}$ and the bottom number corresponds to the person-years $Y_{a,p,c}$ for an age (a), period (p), and birth cohort (c) combination. For example, during 2012 we had $D_{67,2012,1944} = 21$ new cases of breast cancer among women 67 years old who were born in 1944 (triangle in bold) and $D_{67,2012,1945} = 18$ new cases among women of the same age who were born in 1945 (dashed triangle). Corresponding person-years were $Y_{67,2012,1944} = 3663.5$ and $Y_{67,2012,1945} = 3547$, respectively.

five sites for women (breast, colorectal, lung, skin melanoma, and others) and five sites for men (prostate, colorectal, lung, skin melanoma, and others).

Our second source of information was the Swiss Federal Statistical Office (FSO), which produces population figures for each Swiss canton, informing on how many people are alive at the beginning of a calendar year by age and sex.

2.2 | Age, period, and cohort tabulation

Data were aggregated on the three dimensions of age, period, and cohort, using the well-known triangle representation of the Lexis diagram (Figure 1). For each triangle, data include the number of incident cases for the corresponding age a ($a = 0, \dots, 98, 99+$), period p ($p = 1982, \dots, 2016$), and birth cohort c ($c = 1882, \dots, 2016$) combination ($D_{a,p,c}$), and the number of person-years for the same combination ($Y_{a,p,c}$) in the general population, the latter being calculated from population data using the classical method presented in Carstensen (2007). APC-specific incidence rates are then defined by: $\lambda_{a,p,c} = D_{a,p,c}/Y_{a,p,c}$. Each ratio $\lambda_{a,p,c}$ in the Lexis diagram acts as an "observation" in most models considered (see Section 4). Therefore, for each sex and cancer site, we worked with 7000 observations obtained by combining the following factors: 35 periods ($p = 1982, \dots, 2016$), 100 ages ($a = 0, \dots, 98, 99+$), and 2 triangles (upper/lower) separating people in two different birth cohorts (e.g., a person reaching the age of 67 in 2012 may have been born in either 1945 or 1944; see Figure 1). Some of the models below consider only the age and period tabulation:

$$\lambda_{a,p} = \frac{1}{2}\left(\frac{D_{a,p,c}}{Y_{a,p,c}} + \frac{D_{a,p,c+1}}{Y_{a,p,c+1}}\right) = \frac{D_{a,p}}{Y_{a,p}} \text{ (with } c = p - a), \tag{1}$$

or only the age and cohort tabulation:

$$\lambda_{a,c} = \frac{1}{2}\left(\frac{D_{a,p,c}}{Y_{a,p,c}} + \frac{D_{a,p+1,c}}{Y_{a,p+1,c}}\right) = \frac{D_{a,c}}{Y_{a,c}} \text{ (with } p = c + a). \tag{2}$$

TABLE 1 Summary of quantities related to the triangle tabulation of the Lexis diagram.

a	Age
p	Period
c	Cohort
$D_{a,p}$ ($D_{a,c}$; $D_{a,p,c}$)	Number of new cancer cases for age a and period p (respectively, age a and period c; or age a, period p, and cohort c)
$Y_{a,p}$ ($Y_{a,c}$; $Y_{a,p,c}$)	Person-years for age a and period p (respectively, age a and period c; or age a, period p, and cohort c)
$\boldsymbol{\lambda}_{a,p}$ ($\boldsymbol{\lambda}_{a,c}$; $\boldsymbol{\lambda}_{a,p,c}$)	Incidence rates for age a and period p (respectively, age a and period c; or age a, period p, and cohort c)
Y_a^*	Person-years for age a of a reference population
$\boldsymbol{\lambda}_p^*$	Standardized incidence rates of year p

2.3 | Standardized incidence rates

To compare incidence rates over periods using a single quantity, age-specific incidence rates should be combined taking into account changes in the population structure over time. This can be achieved by weighting the age-specific incidence rates of a period p ($\lambda_{a,p}$) for the age distribution of a reference population Y_a^* ($a = 0, \dots, 99+$), resulting in so-called *standardized incident rates*:

$$\lambda_p^* = \frac{\sum_{a=0}^{99+} \lambda_{a,p} Y_a^*}{\sum_{a=0}^{99+} Y_a^*}. \tag{3}$$

It has been shown (Spiegelman & Marks, 1966) that when comparing incidence rates over time, the choice of the reference population has little impact on the results. The same is true when making predictions. In this study, we chose the population of the year 2000 of the aggregated cantons of Vaud, Geneva, and Neuchatel as the reference for standardization. A summary of the notations used for quantities related to the Lexis diagram is provided in Table 1.

The standardized incidence rates λ_p^* ($p = 1982, \dots, 2016$) can be seen in Figure 2 for each sex and cancer site. While some rates show a monotonic trend, for example, lung cancer, which continues to increase for women and decrease for men at the present time, other rates show a recent stabilization (breast cancer, skin melanoma) or a trend reversal, in particular prostate cancer, whose incidence began to decrease in the 2000s after decades of increase.

While the majority of the models considered in this paper use tabulated data by age, period, and/or cohort, predicting specific incidence rates $\lambda_{a,p}$, $\lambda_{a,c}$ or $\lambda_{a,p,c}$, we followed the literature (H. S. Chen et al., 2012; Clements, 2005; Møller et al., 2003) and compared their prediction performance using the standardized incidence rates λ_p^* (3). This is the most widely used choice when predicting the cancer burden (Møller et al., 2002; Rapiti et al., 2014) and allows the comparison of the predictions by ARIMA and simple linear regression models which model directly standardized incidence rates (see Section 4).

3 | EVALUATING PREDICTION PERFORMANCE

3.1 | Leave-future-out cross-validation

Models were compared by repeatedly applying the principle of leave-future-out cross-validation (Bürkner et al., 2020). The procedure can be summarized as follows. *First*, for a given sex/cancer site combination, chose a cutoff time t in the range: $t = 2001, \dots, 2015$. Given the cutoff, the leave-future-out cross-validation consists in (a) fitting a model on incidence rates from $T_0 = 1982$ until year t (*training set*), (b) predicting incidence rates for the second part of the data (*test set*), that is, from the year $(t + 1)$ until the last year available $T = 2016$, and (c) evaluating the prediction performance from $(t + 1)$ to T according to criteria below. *Second*, repeat the procedure moving t in the range $t = 2001, \dots, 2015$ and for each sex/cancer site combination. The range for the cutoff t is chosen to ensure having at least 20 observations in the training set and at least one observation in the test set.

Because we had five cancer sites for each sex and 15 possible cutoffs for each sex/cancer site combination ($t = 2001, \dots, 2015$), we obtained 150 different scenarios (see Figure 3). All models presented in Section 4 will be fitted on each of these 150 scenarios.

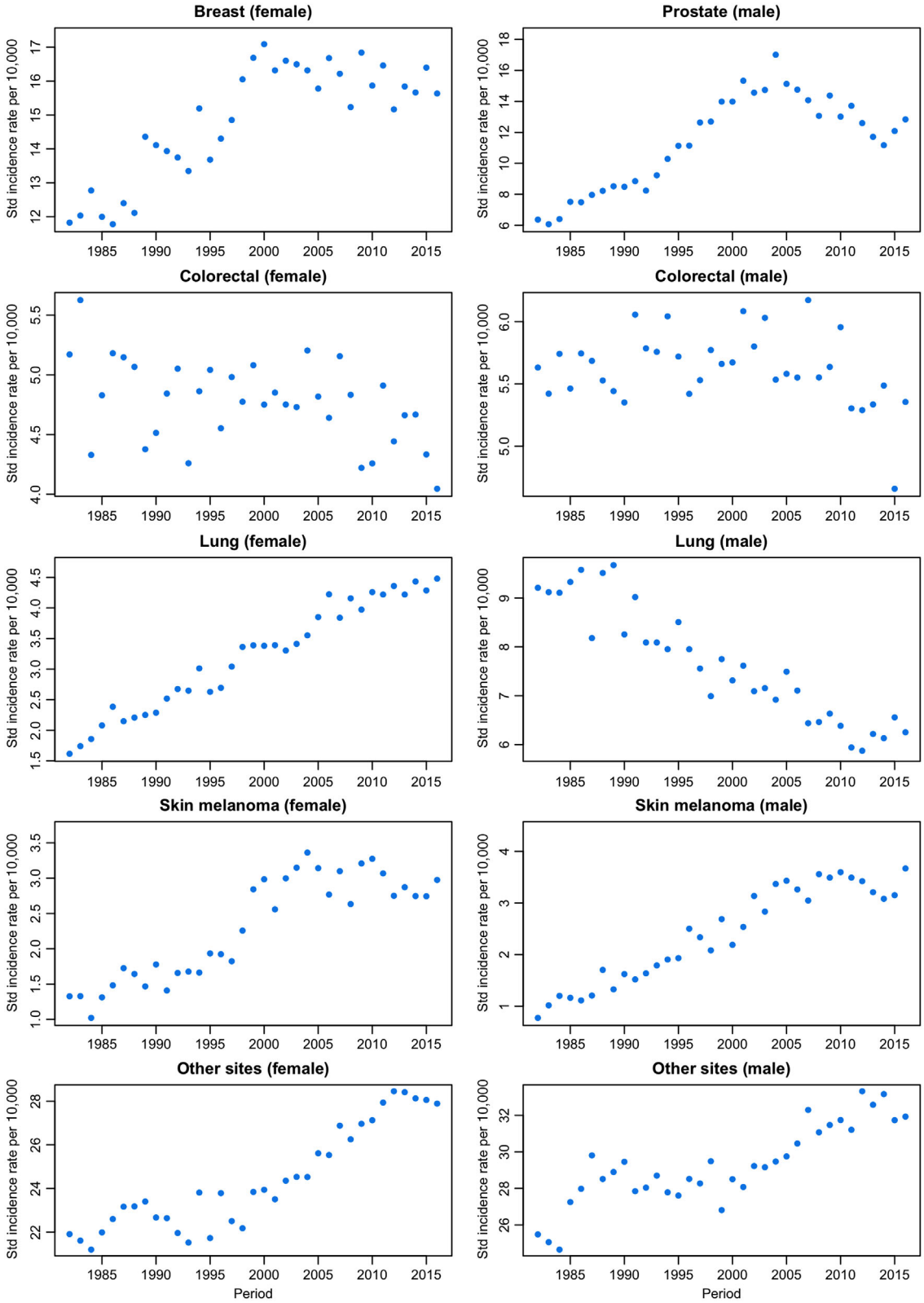

FIGURE 2 Standardized incidence rates per 10,000 inhabitants over the period 1982–2016 for the most common cancer sites and the two sexes. Combined data from the Swiss registries of Vaud, Geneva, and Neuchatel. Reference for standardization: combined population of Vaud, Geneva, and Neuchatel in 2000.

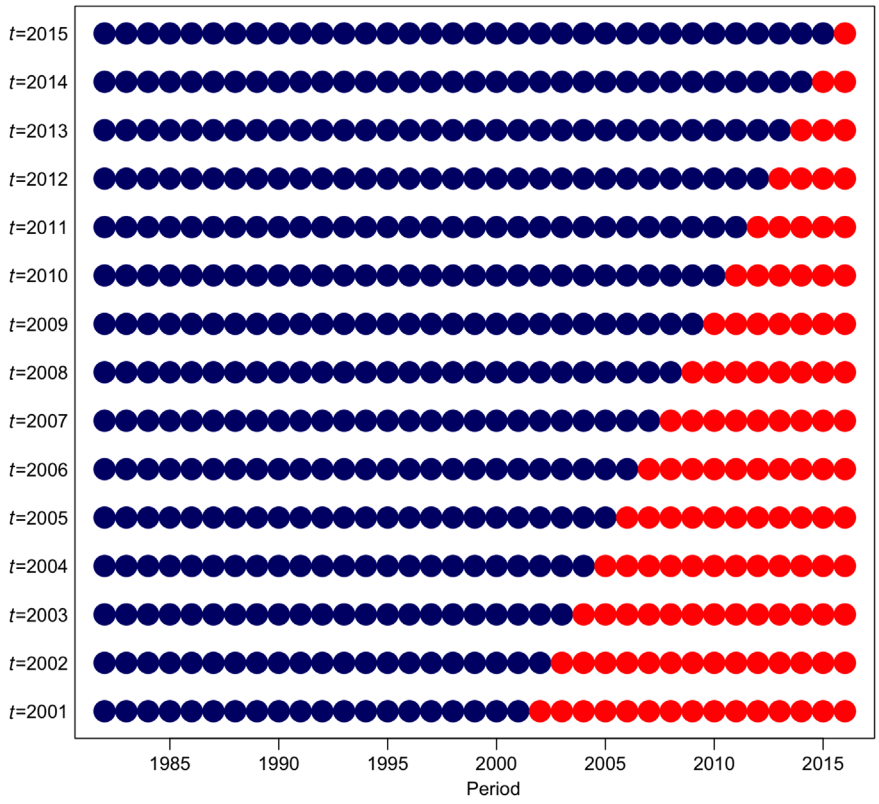

FIGURE 3 Illustration of the 15 leave-future out cross-validation scenarios for each sex and cancer site, based on partitioning the time range into a training set (blue dots) and a test set (red dots) according to a cutoff time t (working with five cancer sites for women and five for men resulted in 150 different scenarios).

3.2 | Comparison criteria

Our main criterion of prediction performance was the normalized root mean squared error (NRMSE) (Hyndman & Koehler, 2006), a measure of point prediction accuracy. This criterion takes the following form in each scenario, that is, for a given cutoff time t, $t = 2001, \ldots, 2015$, and for each sex/cancer site combination:

$$\text{NRMSE}(t) = \frac{\sqrt{\sum_{x=t+1}^{T} \left(\hat{\lambda}_x^* - \lambda_x^*\right)^2 / (T-t)}}{\sum_{x=t+1}^{T} \lambda_x^* / (T-t)}. \tag{4}$$

Here λ_x^* is the observed (i.e., actual) standardized incidence rates for the period x in the test set and for the considered sex/cancer site combination, and $\hat{\lambda}_x^*$ the corresponding predicted standardized incidence rate. The denominator of (4) represents the mean standardized incidence rate in the test set, allowing normalizing of the statistics. The prediction performance (prediction accuracy) of each model was evaluated by the mean NRMSE (M-NRMSE) across the 150 scenarios. A smaller M-NRMSE value indicates better accuracy, with a minimum of 0. We also looked at three time horizons for prediction, assessing separately the short-term M-NRMSE (1–5 years, $t+1 \leq x \leq t+5$), medium-term M-NRMSE (6–10 years, $t+6 \leq x \leq t+10$), and long-term M-NRMSE (11–15 years, $x \geq t+11$). This represented 150, 100, and 50 scenarios, respectively.

In a first sensitivity analysis, we considered an alternative of (4), the normalized mean absolute error (NMAE), defined as follows for a given cutoff time, $t = 2001, \ldots, 2015$, and for each sex/cancer site combination (Hyndman & Koehler, 2006):

$$\text{NMAE}(t) = \frac{\sum_{x=t+1}^{T} |\hat{\lambda}_x^* - \lambda_x^*| / (T-t)}{\sum_{x=t+1}^{T} \lambda_x^* / (T-t)}, \tag{5}$$

and considered the mean NMAE (M-NMAE) across the 150 scenarios. In a second sensitivity analysis, we considered the median (instead of the mean) of the NRMSE and the NMAE across scenarios (Med-NRMSE and Med-NMAE).

As an additional comparison criterion, we evaluated the quality of the 95% prediction intervals provided by the different methods to inform on the prediction uncertainty. For this, we used the coverage rate (CR) and the interval score (IS) (Gneiting et al., 2007). The CR measures the empirical coverage of a prediction interval, that is, the probability that it contains the actual standardized incidence rate and is defined as follows for a given cutoff time $t = 2001, \dots, 2015$, and for each sex/cancer site combination:

$$CR(t) = \frac{\sum_{x=t+1}^{T} I(L_x < \lambda_x^* < U_x)}{(T-t)}, \tag{6}$$

where L_x and U_x are the lower and upper bounds of a 95% prediction interval calculated for period x ($x = t+1, \dots, T$) and I is an indicator function. The IS is a strictly proper scoring rule, which is related to the width of a prediction interval, with a penalty in cases where the interval does not contain the actual standardized incidence rate, and is defined as follows for a given cutoff time $t = 2001, \dots, 2015$, and for each sex/cancer site combination (Gneiting et al., 2007):

$$IS(t) = \frac{\sum_{x=t+1}^{T} \left\{ (U_x - L_x) + \frac{2}{0.05} \left[(L_x - \lambda_x^*) I(\lambda_x^* < L_x) + (\lambda_x^* - U_x) I(\lambda_x^* > U_x) \right] \right\}}{(T-t)}. \tag{7}$$

The CR of a prediction interval should be as close to 95% as possible, while the IS should be as small as possible with a minimum of 0. As for our primary criteria, CR and IS were averaged across the 150 scenarios, obtaining a mean CR and a mean IS (M-CR and M-IS).

Finally, we reported the percentage of scenarios for which the fitting algorithms used by the different methods converged (i.e., they were able to produce predictions), which was not always 100%. Note that all the above criteria were calculated excluding the few cases where the computation did not converge.

All models below will be evaluated and compared according to all criteria (M-NRMSE, med-NMRSE, M-NMAE, Med-NMAE, M-CR, M-IS, and convergence rate). All detailed results are given in the Appendix of the Supporting Information.

4 | PREDICTION METHODS

4.1 | Generalized linear models

We considered the Poisson GLM as presented in Chapter 2.4 of A. C. Cameron and Trivedi (2013a). The number of cases $D_{a,p}$ ($D_{a,c}$) in a certain area of the Lexis diagram (see Figure 1) is assumed to follow a Poisson distribution whose mean $\lambda_{a,p} \cdot Y_{a,p}$ ($\lambda_{a,c} \cdot Y_{a,c}$) depends (usually via a log link) on the age and period (or age and cohort). The person-years $Y_{a,p}$ ($Y_{a,c}$) is an offset of the model. Due to the *age = period−cohort* relationship, the three variables cannot be introduced simultaneously in a model, so that we considered data either in squares of the Lexis diagram (age and period) or in parallelograms (age and cohort, see Figure 1), that is, only two effects were considered at a time, with a possible interaction between the two:

$$\log\left(\frac{D_{a,z}}{Y_{a,z}}\right) = \log\left(\lambda_{a,z}\right) = ns_k(a) + f(z) + g(a) \cdot h(z) \tag{8}$$

with $z = \{p, c\}$; $f, g, h = \{\emptyset, id, ns_k\}; k = 1, \dots, 4$. The age a was introduced into the model with natural splines ns_k, that is, cubic splines based on k knots, linear outside the extreme knots (Ruppert et al., 2003). The period (or the cohort) was either absent ($f = \emptyset$), either introduced linearly, that is, via an identity function ($f = id$) or by splines ($f = ns_k$). Both age and period (or age and cohort) were introduced either linearly or by splines in the interaction term, with the constraint that the degrees of freedom for the period (or the cohort) effect must be lower or equal in the interaction term than in the main effect. So, if the period (or the cohort) was introduced linearly as the main effect ($f = id$), it was not introduced via splines in the interaction ($h \neq ns_k$), and if the period (or the cohort) was not introduced as a main effect ($f = \emptyset$), it was not in an interaction term ($h = \emptyset$). For a given number of knots k, this defined five models without interaction and 12 models with interaction. Varying the number of knots between $k = 1$ and $k = 4$ (beyond which we encountered problems of lack of convergence of the fitting algorithm), we obtained $4 \times (5 + 12) = 68$ GLM. The Lee–Carter model (1992), a well-

established model used in demography to predict mortality rates, is one of those 68 models where the age and the period are introduced with splines, with an interaction between the two.

Finally, we considered a GLM obtained via a model selection strategy that consists in selecting among the 68 GLM models the one that best fits the data in the training set according to the Akaike information criterion (AIC), so that a possibly different GLM is selected in each scenario. This corresponds to a 69th GLM that we call GLM AIC.

The well-known joinpoint regression (Lerman, 1980) was included as a 70th GLM. This is an age-period model with the period introduced with linear splines, with knots representing times where a change of trend occurs. Unlike classical splines, where the knots are chosen a priori, knots in joinpoint regression are determined from the data according to AIC. We only considered a single knot joinpoint model.

Predictions of incidence rates were obtained by extrapolating period, respectively, cohort effects. For age-cohort models, cohort effects must be extrapolated for future periods, with the particularity of requiring the prediction of the effect of some new cohorts and the removal of the effect of old cohorts, as each year some cohorts exit and new cohorts enter. Predictions of $\lambda_{a,c}$ or $\lambda_{a,p}$ ($\hat{\lambda}_{a,c}$ or $\hat{\lambda}_{a,p}$) were then aggregated and standardized according to (3) to obtain predicted standardized incidence rates $\hat{\lambda}^*_p$. The variance of $\hat{\lambda}_{a,c}$ (or $\hat{\lambda}_{a,p}$) was estimated by the delta method as in Chapter 3 of A. C. Cameron and Trivedi (2013b), and prediction intervals were obtained based on this variance following the implementation provided in the `R package trending` (Schumacher & Jombart, 2021).

4.2 | Age-period-cohort models

APC models, introduced by Holford (1983), are Poisson GLM allowing to include simultaneously age, period, and cohort effects. In order to make the model identifiable despite the linear relationship between the three variables, some constraints are added (Carstensen, 2007). Whereas age, period, and cohort were originally considered as factors, forcing a coarse tabulation of 5-year groups (Clayton & Schifflers, 1987; Holford, 1983; Møller et al., 2002), more recent developments in APC, which we followed in our study, adopted splines to model three effects (Carstensen, 2007; Carstensen et al., 2022). With the canonical log link of Poisson GLM, a spline APC model takes the following form:

$$\log\left(\lambda_{a,p,c}\right) = ns_k\left(a\right) + ns_k\left(p\right) + ns_k\left(c\right) + \text{drift}. \tag{9}$$

Here the drift is a linear trend effect, ascribed to both period and cohort and $ns_k(p)$ (respectively, $ns_k(c)$) are residual nonlinear effects specific to the period (respectively, the cohort). For these splines, we considered between k =3, 4, or 5 knots. In addition to the canonical log link, other link functions (Dunn & Smyth, 2018) have been compared in the literature for their prediction performance, where the 1/5 power link performed best (Møller et al., 2003). For this reason, we have adopted the 1/5 power link in our comparisons, as an alternative to the log link (9).

Two options have been adopted in the literature to extrapolate the incidence rates $\hat{\lambda}_{a,p,c}$ for future periods and cohorts. The most common one consists in extrapolating only the drift (Yu et al., 2019); the alternative consists in adding a linear extrapolation of the nonlinear effects (Rutherford et al., 2012). Both options are compared in our study. In what follows, the two extrapolation strategies are referred to as *drift only*, and *all effects*, respectively. Varying the number of knots between k =3, 4 or 5, and considering two link functions and two extrapolation strategies, we obtained $3 \times 2 \times 2 = 12$ APC models.

As with GLM, the predictions $\hat{\lambda}_{a,p,c}$ were finally aggregated and standardized as in (3) to produce predicted standardized incidence rates $\hat{\lambda}^*_p$ and prediction intervals.

4.3 | Bayesian age-period-cohort models

A BAPC model (Riebler et al., 2012; Riebler & Held, 2017; Schmid & Held, 2004) has the same formulation as an APC model, the difference being how age, period, and cohort effects are included in the model and how they are fitted. Instead of using splines, Bayesian models are based on the first- or second-order random walk (RW_1 or RW_2) specification for each effect of the model:

$$\log\left(\lambda_{a,p,c}\right) = RW_l\left(\text{a}\right) + RW_l\left(p\right) + RW_l\left(c\right), \tag{10}$$

where $l = 1, 2$ is the order of the random walk. Considering, for instance, the period effect (the same holds for age and cohort effects), a Bayesian first-order random walk takes the following form:

$$RW_1(p) = \Delta p \sim N\left(0, \sigma_1^2\right), \tag{11}$$

where Δp is the difference of the last two period effects. With the RW_1 model, each (age, period, and cohort) effect is taken as being equal to the previous one plus a random number from a normal distribution centered on zero, with variance σ_1^2. The latter acts as a smoothing parameter: with an infinite variance, the fitted model will pass through all the data points in the training set, while a small prior variance will give rise to a smoother fit. We considered different choices for the hyper-prior distribution of parameter σ_1^2. Our first option was the most commonly adopted distribution, that is, a flat (noninformative) gamma(1, 5e−5) for the three effects of age, period, and cohort (Riebler & Held, 2017). As alternative options, we considered a tighter gamma(1, 5e−3), a larger gamma(1,5e−7), and PC(u=1,α=0.01) (Lindgren & Rue, 2015) priors for the three effects, as well as an option with a gamma(1, 9e−4) for the age effect and gamma(1, 2.5e−4) for the period and cohort effects, as in Riebler and Held (2017). Considering that we have 7000 observations in the Lexis diagram, the hyper-prior has a limited effect. When making predictions from a Bayesian RW_1 model, the last fitted period and cohort effects are extrapolated as constant.

Considering again the period effect as an example, a Bayesian second-order random walk takes the form:

$$RW_2(p) = \Delta^2 p \sim N\left(0, \sigma_2^2\right), \tag{12}$$

where $\Delta^2 p$ is the second difference (difference of the difference) involving the last three period effects. In a RW_2 model, each age, period, and cohort effect is based on the two previous effects and adding a random number from a normal distribution centered on zero, with variance σ_2^2. The latter acts again as a smoothing parameter: the smaller the prior variance, the smoother the fit. The same five priors adopted for σ_1^2 were adopted for σ_2^2. In a Bayesian RW_2 model, the future period and cohort effects are predicted by a linear trend based on the last two fitted effects. Compared to an RW_1 model, an RW_2 model will be generally smoother.

Since we considered five priors for each of the RW_1 and RW_2 models, we included a total of 10 BAPC in our comparison. As with GLM and APC models, the predicted rates $\hat{\lambda}_{a,p,c}$ from a BAPC model should be combined to obtain predictions of standardized incidence rates $\hat{\lambda}_p^*$. Bayesian RW_1 and RW_2 models were fitted using the `INLA` (Lindgren & Rue, 2015) and `BAPC packages` (Riebler & Held, 2017) from `R` software (R Core Team, 2022) to get point predictions and prediction intervals, respectively. The `package INLA` is based on the works of Rue (2009), Martins (2013), and Lindgren (2011, 2008, 2015).

4.4 | ARIMA time series models

ARIMA time series models are econometric models defined by combining a difference autoregressive model with a moving average model (Hamilton, 1994). Let $\Delta^d \lambda_p^*$ be the standardized rates λ_p^* differenced d times. The ARIMA(h,d,q) is expressed as

$$\Delta^d \lambda_p^* = \alpha_0 + \alpha_1 \Delta^d \lambda_{p-1}^* + \alpha_2 \Delta^d \lambda_{p-2}^* + \cdots + \alpha_h \Delta^d \lambda_{p-h}^* + \epsilon_p + \theta_1 \epsilon_{p-1} + \theta_2 \epsilon_{p-2} + \cdots + \theta_q \epsilon_{p-q}, \tag{13}$$

where ϵ_p are normally distributed residuals, $\alpha_1,\ldots,\alpha_h$ are the coefficients of the autoregressive (AR) part of the model, $\theta_1, \ldots, \theta_q$ are the coefficients of the moving average (MA) part, and α_0 is a constant. In an ARIMA(h,d,q) the predictors are lagged h data points for the autoregressive part and q residuals are considered for the moving average part, which are all d differenced. We considered all orders from (0,0,0) to (3,3,3) for (h,d,q) which represents $4^3 = 64$ different models. As for GLM, we finally considered an ARIMA obtained via a model selection strategy that consists in selecting among the 64 ARIMA models the one that best fits the data in the training set according to the AIC, so that a possibly different ARIMA is selected in each scenario. This corresponds to the 65th ARIMA method that we call ARIMA AIC.

Unlike the models presented in the previous sections, the predictions obtained with an ARIMA model are based directly on an extrapolation of the standardized rates (without the need for any aggregation). We used the function ARIMA from R (R Core Team, 2022) to fit the models, predict the incidence rates, and compute the prediction intervals.

4.5 | Linear models

In order to get another standard comparator for the above methods, we finally considered a simple linear model (LM) fitted on the last r data points of the training set:

$$\lambda_p^* = \alpha + \beta p + \epsilon_p, \tag{14}$$

where $p = (t - r + 1), \dots, t$, ϵ_p are normally distributed residuals and α and β are the intercept and slope of the model, respectively. Letting r vary between 3 and 10 data points, we considered eight LM. In these models, predictions of standardized incidence rates are made by extrapolating the fitted trend, and prediction intervals are obtained in the classical framework of regression models.

In summary, considering the five classes of models described in this section, we compared 165 different models: 70 GLM, 12 APC, 10 BAPC, 65 ARIMA, and 8 LM. An exhaustive list of all the models compared can be found in Figures A1–A4 of the Supporting Information Appendix.

5 | RESULTS

All models described in Section 4 were fitted to the 150 available scenarios detailed in Section 3.1 and evaluated according to the performance criteria presented in Section 3.2. For each model, the NRMSE and IS distribution over the 150 scenarios is given in Figures A1–A8 of the Supporting Information Appendix. Most of these distributions show large variability and are highly skewed, with M-NRMSE > Med-NRMSE, as M-NRMSE strongly penalizes the poor performance of some models in particularly difficult scenarios, for example, in the case of a trend reversal (and the same is true for the M-IS).

5.1 | Illustration of two selected scenarios

To illustrate the heterogeneity of predictions among models, we have plotted in Figure 4 two selected scenarios, one for female lung cancer and one for (male) prostate cancer, both with a cutoff time of $t = 2010$. In the former, incidence rates are always increasing, representing a simple setting for predictions, while in the latter the incidences begin to decrease around 2005, giving a more challenging situation for predictions. It can be seen that in the former scenario almost all models predict similar incidence rates, whereas in the latter the heterogeneity among predictions is impressive, as a trend reversal occurs towards the end of the training set. In this setting, some models adapt to the new trend (e.g., an APC extrapolating *all effects*, represented in green), while others fail to adapt (e.g., an APC extrapolating the *drift only*, represented in dark cyan). We can observe that highly flexible models, although fitting closely to the data points in the training set, are not necessarily the best for prediction in the test set (e.g., ARIMA (3,3,0) in red).

The Supporting Information Appendix contains more detailed results for prostate cancer. Figure A9 shows the predictions of the different models for each of the 15 scenarios (i.e., varying the cutoff time t) for this cancer site. When the cutoff time occurs before the trend reversal, all models predict excessively high incidence rates, but some models adapt quicker than others when the cutoff time occurs during or just after the trend reversal. For each model, Figures A10–A13 give the NMRSE distribution over the 15 scenarios for prostate cancer, showing a generally worse performance of the methods (higher NMRSE values and more skewed distribution) than what we had in Figures A1–A4 (when considering all cancer sites). The NRMSE and IS distributions of each model over the 15 scenarios for all 10 sites considered in this study can be found in the Supporting Information. In what follows, we discuss and compare the prediction performance achieved by the different models.

5.2 | Comparison within classes of models

In the class of GLM (Section 4.1), recall that we varied (a) the number of knots k of the splines between 1 and 4, and (b) the variables included in the model and the functional forms used. Models with $k = 3$ and $k = 4$ knots achieved smaller M-NMRSE (Figure 5a). Models including age and period performed better than models including only age or age and cohort, with better performance when the period was included via splines rather than just linear, while models including

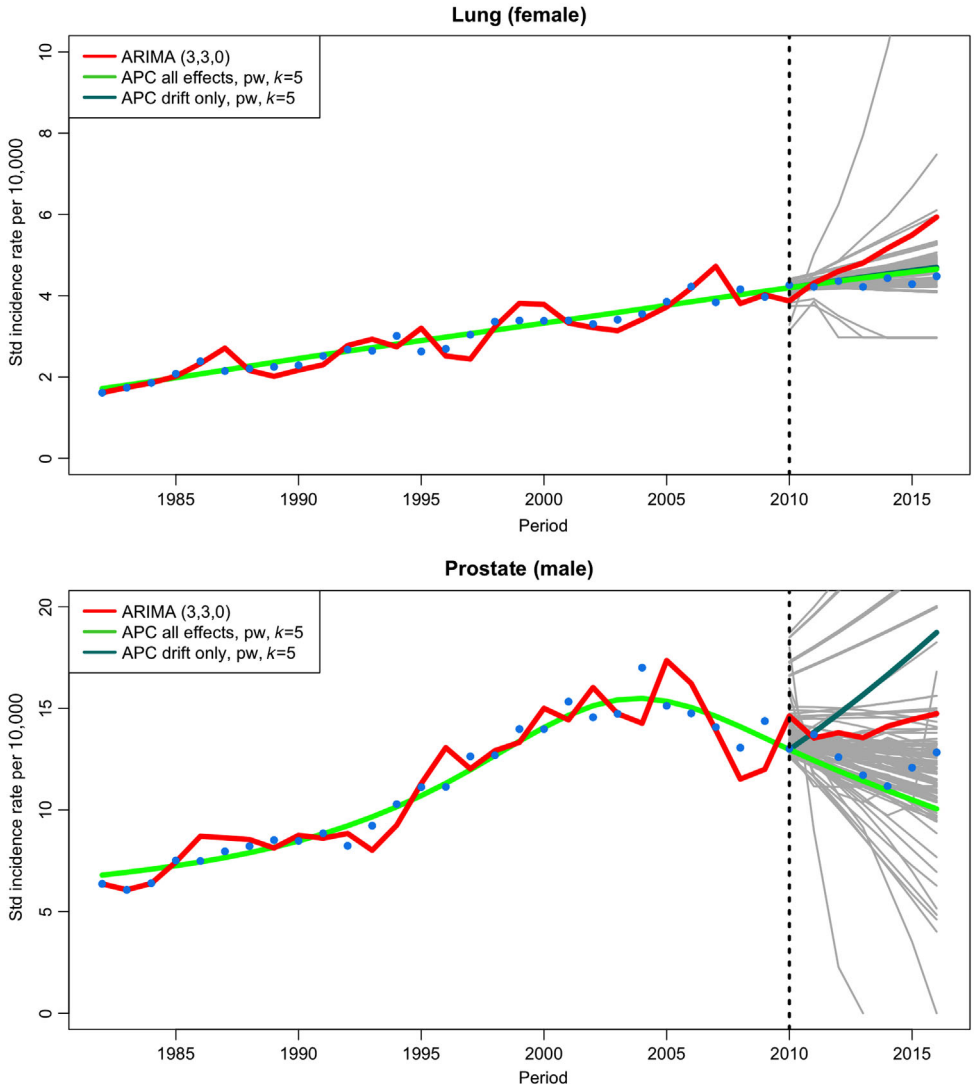

FIGURE 4 Illustration of predictions made by the different models in one of the 15 leave-future-out cross-validation scenarios (training set 1982–2010; test set 2011–2016) for lung (female) and prostate (male) cancer standardized incidence rates. Combined data from the Swiss registries of Vaud, Geneva, and Neuchatel, 1982–2016. Three models are highlighted (APC extrapolating only the drift, APC extrapolating the drift + all nonlinear effects, and ARIMA (3,3,0)). All other models are in gray. APC, age-period-cohort; ARIMA, autoregressive integrated moving average.

an interaction term performed similarly to those without interaction (Figure 5b). The best GLM with $k = 4$ knots was a model including age and period, along with a linear interaction between the two variables (M-NRMSE = 0.138), while the best GLM with $k = 3$ knots was a model including age and period without interaction (M-NRMSE = 0.143), the latter being simpler and converging more often than the former (100% vs. 96%). Both models performed much better than the more complex Lee–Carter model (M-NRMSE = 0.180). On the other hand, the use of AIC for selecting the GLM showed an extremely poor performance, as did the joinpoint model (both M-NRMSE > 7; see also Figures A1 and A2 in the Supporting Information Appendix).

In the APC class of models (Section 4.2), we varied the number of knots k between 3 and 5, while considering two possible link functions (log or 1/5 power), and two different extrapolation strategies (*drift only* or *all effects*). Results are summarized in Figure 5c. We found that using the 1/5 power link improved the predictions compared to using the canonical log link, especially when extrapolating the drift only. The optimal number of knots was $k = 3$ when using the first extrapolation strategy (*drift only*) and $k = 5$ when using the second (*all effects*). In general, the second extrapolation strategy improved predictions compared to the first. The best APC model was thus a model using the 1/5 power link, with five knots and extrapolating *all effects* (M-NRMSE = 0.140; see also Figure A3 in the Supporting Information Appendix).

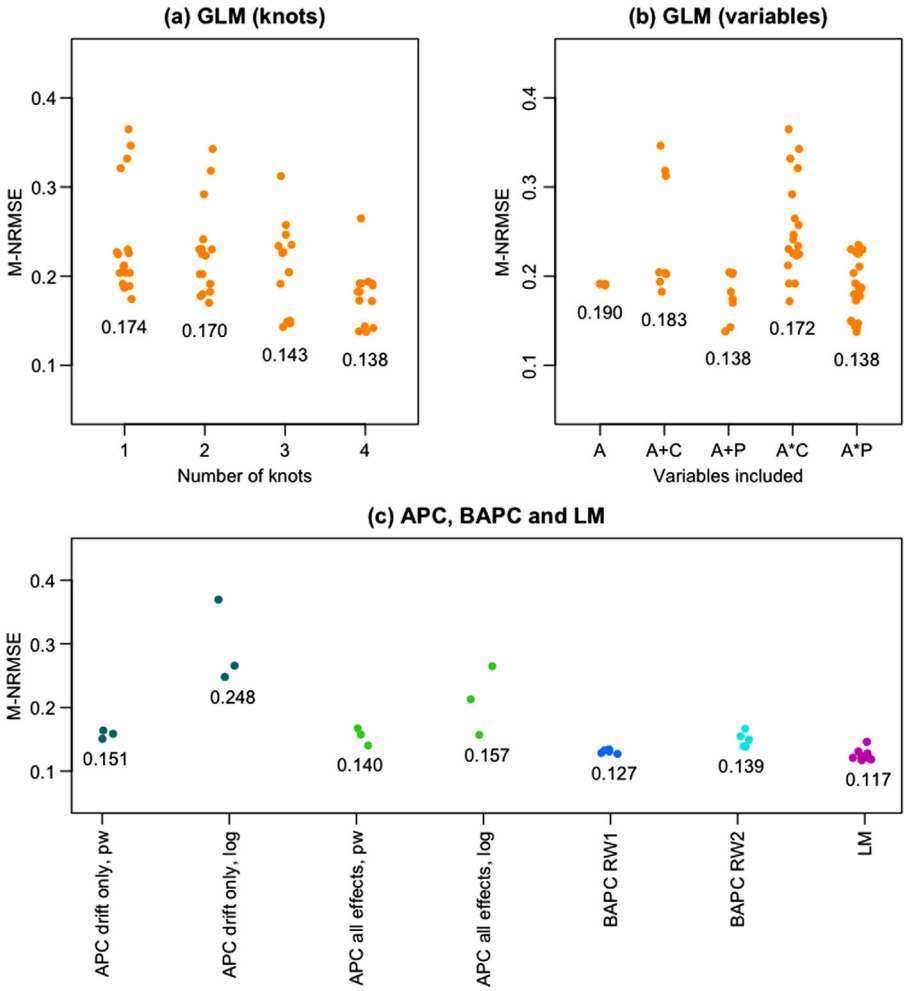

FIGURE 5 Mean normalized root mean square error (M-NRMSE) for GLM (a) and (b); APC, BAPC, and LM (c) across 150 scenarios (15 leave-future-out cross-validation scenarios for each sex and five cancer locations per sex) obtained on combined data from the Swiss registries of Vaud, Geneva, and Neuchatel, 1982–2016. In the labels, "A" indicates GLM only including age; "A+P" (respectively, "A+C") indicates GLM including age and period (respectively, age and cohort) without interaction; "A · P" ("A · C") indicates GLM including age, period (respectively, age, cohort) with interaction. For APC, "drift only" and "all effects" refer to prediction strategies extrapolating only the drift, respectively, the drift + all nonlinear effects, "log" indicates the logarithmic link, and "pw" is the 1/5 power link. For BAPC, RW_1 and RW_2 refer to first, and second-order random walk. APC, age-period-cohort; BAPC, Bayesian age-period-cohort; GLM, generalized linear models; LM, linear model.

In the BAPC class of models (Section 4.3), the choice of the priors for the period, cohort, and age effects had almost no impact on the M-NRMSE, while (using a PC hyper-prior) the performance was better for the RW_1 model (M-NRMSE = 0.127) than for the more complex RW_2 model (M-NRMSE = 0.139), as summarized in Figure 5c (see also Figure A3 in the Supporting Information Appendix).

In the ARIMA class of models (Section 4.4), we varied the autoregressive, integrated, and moving average orders from 0 to 3. Results are summarized in Figure 6. Changing the autoregressive and the moving average order did not have much impact on the prediction performance. On the other hand, working with difference data did improve the performance up to order 1, but got worse beyond. The smallest M-NRMSE was achieved by an ARIMA (2,1,1) (M-NRMSE = 0.078), with a convergence of 96%, while the best ARIMA with a convergence of 100% was ARIMA (1,1,0) (M-NRMSE = 0.084). As for GLM, using the AIC criterion to select a possibly different ARIMA model depending on the scenario resulted in a worse prediction performance (M-NRMSE = 0.119) than when systematically opting for an ARIMA (2,1,1) or (1,1,0) (see also Figure A4 in the Supporting Information Appendix).

Finally, in the class of LM (Section 4.5), the best performance was achieved using the last $r = 7$ data points of the training set to fit the model (M-NRMSE = 0.117; see Figure 5c and Figure A3 of the Supporting Information Appendix).

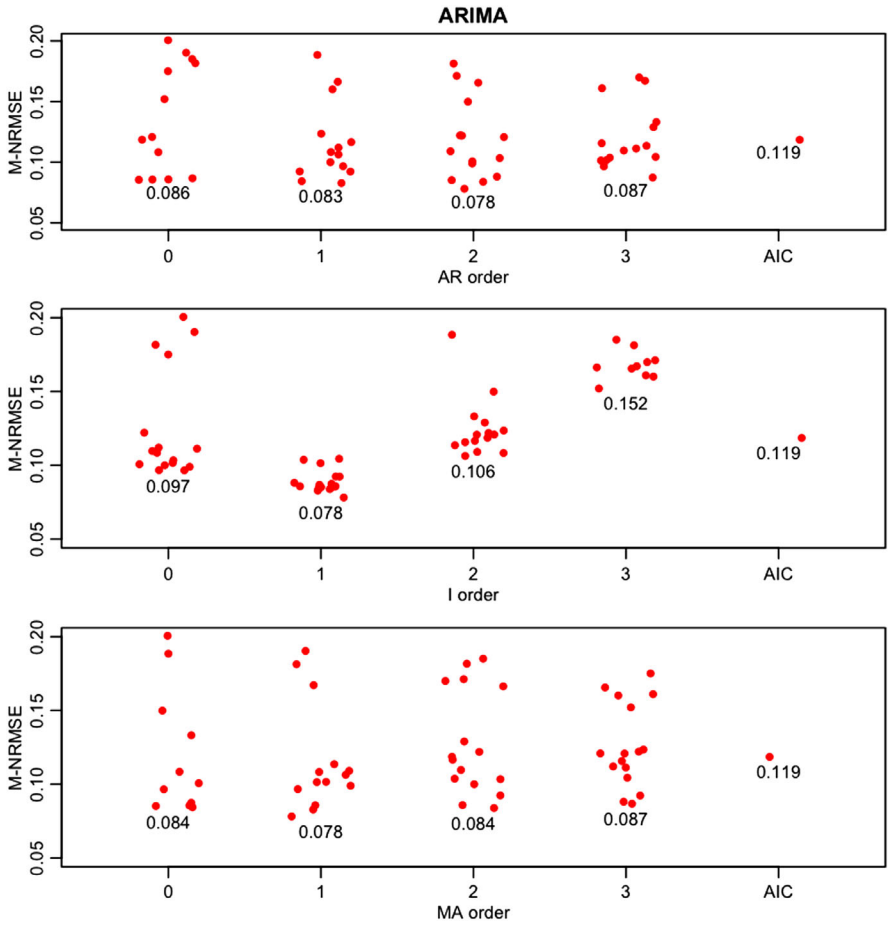

FIGURE 6 Mean normalized root mean square error (M-NRMSE) for ARIMA with orders of each component (AR, I, and MA) between 0 and 3, and an ARIMA AIC across 150 scenarios (15 leave-future-out cross-validation scenarios for each sex and five cancer locations per sex) obtained on combined data from the Swiss registries of Vaud, Geneva, and Neuchatel, 1982–2016. AIC, Akaike Information Criterion; AR, autoregressive; ARIMA, autoregressive integrated moving average; MA, moving average.

The Supporting Information Appendix also contains the same summary results as in Figures 5 and 6 for the M-IS criterion (Figures A14 and A15). Within-class comparisons in terms of M-IS are largely consistent with that obtained using M-NMRSE. The best GLM was a model with three knots including age and period with an interaction (M-IS = 0.199; Figures A14, A5, and A6). The best APC was a model using the 1/5 power link, with five knots and extrapolating *all effects* (M-IS = 0.504; Figures A14 and A7). The best BAPC was a RW_1 model using a gamma(1, 2.5e−4) prior (M-IS = 0.200; Figures A14 and A7). In the ARIMA class, the best performance was achieved by ARIMA(1,1,0) (M-IS = 0.098; Figures A15 and A8), while in the class of LM the best performance was obtained using the last $r = 4$ data points (M-IS = 0.181; Figures A14 and A7).

5.3 | Comparison between classes of models

The performance of the best models from each class identified above based either on M-NRSME or on M-IS, as well as of some other well-known models, is summarized in Table 2, where models are ordered in terms of M-NRSME. Based on this criterion, simple ARIMA models, such as (2,1,1) and (1,1,0) and LM using the last seven data points, performed better than the best models among the more complex (GLM, APC, and BAPC) classes of models. Looking at detailed results in Supporting Information Appendix (Figures A1–A4), one can see that ARIMA models with orders up to (3,1,3) as well as LM models using the last 5–10 data points were also ahead of the other methods. The AIC-based ARIMA outperformed GLM, APC, and BAPC models, while being inferior to simple ARIMA and LM. Finally, the single knot joinpoint and Lee–Carter models showed extremely poor performance.

TABLE 2 Prediction performance for the best models within each class (GLM, APC, BAPC, ARIMA, and LM) according to the mean normalized root mean square error (M-NRMSE) and to the mean interval score (M-IS), and for some other selected models, across 150 scenarios (15 leave-future-out cross-validation scenarios for each sex and five cancer locations per sex) obtained on combined data from the Swiss registries of Vaud, Geneva, and Neuchatel, 1982–2016. Also given is the performance of short-term, medium-term, and long-term predictions provided by M-NRMSE (1–5 years), M-NRMSE (6–10 years), and M-NRMSE (11–15 years). Alternative criteria include the median NMRSE (Med-NRMSE), the mean and median normalized mean absolute error (M-NMAE and Med-NMAE), and the mean coverage rate (M-CR). The last column indicates the percentage of convergence of the methods over the 150 scenarios. Models in this table are ordered according to M-NRMSE.

	M-NRMSE	1–5 years	6–10 years	11–15 years	Med-NRMSE	M-NMAE	Med-NMAE	M-IS	M-CR	Convergence
ARIMA (2,1,1)	0.078	0.068	0.090	0.120	0.078	0.069	0.065	0.103	93.3	96
ARIMA (1,1,0)	0.084	0.071	0.101	0.141	0.079	0.075	0.069	0.098	96.7	100
LM $r = 7$	0.117	0.079	0.153	0.280	0.076	0.103	0.067	0.234	88.8	100
ARIMA AIC	0.119	0.087	0.150	0.223	0.089	0.105	0.079	0.258	78.3	100
BAPC RW1 PC(u=1,α=0.01) prior	0.127	0.085	0.175	0.297	0.083	0.111	0.076	0.205	98.3	89
BAPC RW1 gamma(1,5e-3)	0.129	0.086	0.179	0.303	0.086	0.113	0.074	0.200	99.5	88
LM $r = 4$	0.131	0.091	0.168	0.293	0.083	0.116	0.075	0.181	94.6	100
GLM $ns_4(a) + ns_4(p) + a \cdot p$	0.138	0.087	0.188	0.357	0.088	0.120	0.074	0.377	61.6	96
BAPC RW2 PC prior	0.139	0.084	0.178	0.384	0.091	0.119	0.074	0.337	99.8	86
APC all effects, pw, $k = 5$	0.140	0.088	0.178	0.360	0.093	0.123	0.078	0.504	61.1	97
GLM $ns_3(a) + ns_3(p)$	0.143	0.088	0.186	0.373	0.088	0.124	0.077	0.496	59.8	100
GLM $ns_3(a) + ns_3(p) + ns_3(a) \cdot p$	0.149	0.088	0.194	0.413	0.086	0.128	0.075	0.199	85.5	100
APC drift only, pw $k = 3$	0.151	0.105	0.199	0.303	0.105	0.132	0.094	0.607	51.2	85
GLM Lee Carter, $k = 2$	0.180	0.097	0.233	0.557	0.081	0.154	0.070	0.340	76.7	100
Joinpoint	7.519	0.440	7.982	40.299	0.096	4.327	0.088	11.329	53.7	100

Abbreviations: AIC, Akaike Information Criterion; APC, age-period-cohort; ARIMA, autoregressive integrated moving average; BAPC, Bayesian age-period-cohort; GLM, generalized linear models; LM, linear model. pw, 1/5
power link; RW1, first-order random walk; RW2 second-order random walk.

Largely similar conclusions were obtained in terms of M-IS, with simple models outperforming more complex ones (Table 2 and Figures A5–A8). Again, the best performance was achieved for simple ARIMA models, followed by LM (using the last four points). These simple models also presented an M-CR fairly close to 95%. BAPC models had in general wider prediction intervals, resulting also in too high coverage rates (M-CR close to 100% instead of 95% for some of them). In contrast, coverage rates were clearly too low for GLM and APC, where M-CR could even reach values as low as 50%.

Figure 7 compares the M-NRMSE between model classes separately for each sex and cancer site. Prostate cancer is the site with the largest variability in model accuracy, as incidence rates showed a trend reversal (Figures 4 and A5), followed by skin melanoma for both sexes, because of a rate stabilization in recent years (Figure 2). The APC, BAPC, and GLM models performed particularly poorly in at least one of these difficult cases, while the ARIMA models were among the best methods whatever the cancer site. Here also, similar results were obtained in terms of M-IS (Figure A16).

When evaluating separately the M-NRMSE of the short-term (1–5 years), medium-term (6–10 years), and long-term (11–15 years) predictions, we found that, despite inevitably less good performances for long-term predictions than for mid- and especially short-term predictions, the ranking of the models was largely the same in all three settings (Table 2). Results of our sensitivity analyses (Section 3.2) are also summarized in Table 2. In a first sensitivity analysis, we repeated all calculations using M-NMAE instead of M-NRMSE. The ranking of the models stayed mostly the same. In a second

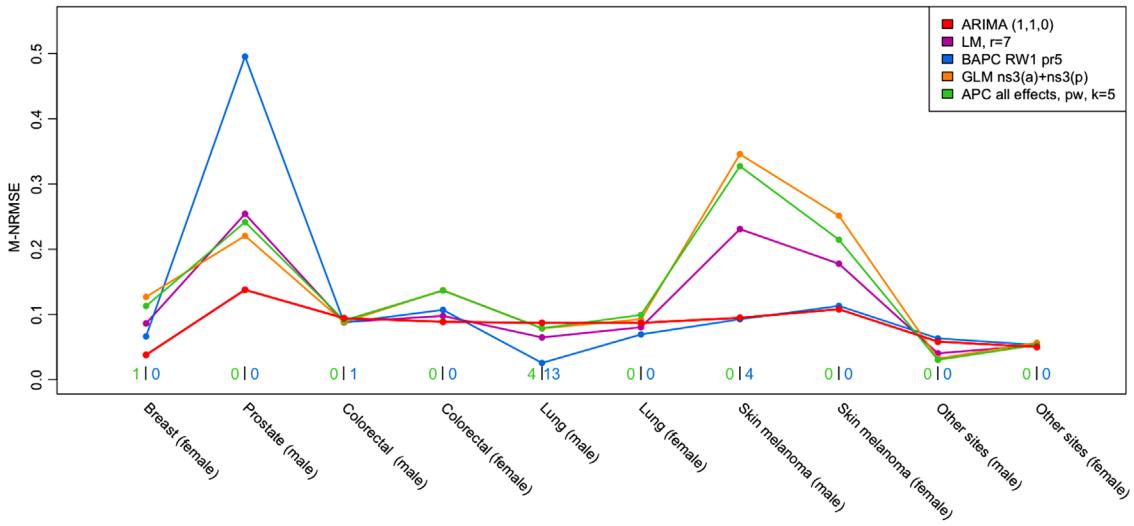

FIGURE 7 Mean normalized root mean square error (M-NRMSE) for one model per class, by sex and cancer site (M-NRSME across 15 leave-future-out cross-validation scenarios on combined data from the Swiss registries of Vaud, Geneva, and Neuchatel, 1982–2016 per sex and cancer site). The number of scenarios for which the fitting algorithm of a method did not converge is indicated at the bottom of the graph for APC|BAPC. The other methods shown always converged. APC, age-period-cohort; ARIMA, autoregressive integrated moving average; BAPC, Bayesian age-period-cohort; GLM, generalized linear models; LM, linear model.

sensitivity analysis, we looked at the Med-NRMSE and Med-NMAE. While the performance of the different models was much closer to each other, the best models remained the simple ARIMA models.

The last column of Table 2 provides the percentage of scenarios for which the fitting algorithm of the different methods converged, which was 100% for LM, above 95% for ARIMA and GLM, and slightly below (down to 85%) for the APC and BAPC models. The computational time of Bayesian and non-Bayesian methods is available in the Supporting Information Appendix (Table A1).

6 | DISCUSSION

The objective of this paper was to compare the performance of statistical models for predicting annual age-standardized cancer incidence rates. The APC method was compared with its Bayesian counterpart BAPC, with simpler GLM approaches including only age and period or age and cohort in the model, and with ARIMA models based only on time series of incidence rates. A simple linear regression model (LM) extrapolating the trend fitted on the last few observed incidence rates was included in the comparison. The prediction performance of each method was evaluated in terms of point prediction accuracy, the quality of prediction intervals, and the convergence rate of the fitting algorithm on a large set of scenarios obtained by leave-future-out cross-validation using real data from Swiss cancer registries.

In the class of GLM, including period led to better prediction performance than including cohort (in addition to age) in the model. Including cohort in the model involves the prediction of new cohort effects mainly based on the most recent ones containing only few observations. These models are therefore more prone to overfitting than those using period effects concerning the whole population. On the other hand, the inclusion of a simple (linear) interaction term for age and period did not improve prediction performance, while including complex interactions (splines) actually worsened predictions due to overfitting. This was also the case for the Lee–Carter and joinpoint models, and for a GLM model selected via the AIC. Prediction intervals were particularly poor for GLM as previously found by another study (Møller et al., 2005), who estimated (as we did) coverage rates of about 50%. The APC approach, although considerably improved by the use of a power link (Møller et al., 2003), performed poorly when predictions were made by extrapolating the drift alone. The drift alone is, however, the most widely used strategy to date (Yu et al., 2019). Further extrapolating all nonlinear effects (Rutherford et al., 2012) improved the predictions in case of a trend reversal. The Bayesian APC models based on second-order random walk performed very similarly to the classical APC models with splines, while a simpler BAPC based on first-order random walk performed slightly better despite less smooth age, period, and cohort effects. For ARIMA, models with low-order autoregressive, integrated, and moving average components are to be preferred to more complex model

structures. A strategy based on selecting the best ARIMA model via AIC overfitted the data and gave worse predictions than systematically choosing a simple structure. Finally, increasing the number of data points to fit a simple linear regression model improved the prediction when including up to 7 years in the past, worsening after this lag.

Comparing models among classes, we obtained the best performance for simple ARIMA models such as (2,1,1) or (1,1,0). Nevertheless, most of the simplest ARIMA models, including component orders up to (3,1,3), were ahead of the other methods. The second-best performing class of models was simple linear regression, where all alternatives using 5–10 data points to fit the trend outperformed the more complex GLM, APC, and BAPC models. Similar results in terms of ranking of models were obtained for short, medium, and long-term predictions, and stratifying by cancer site, despite the less good performance for long-term versus short-term predictions, and for cancers showing a trend reversal. Considering absolute errors (NMAE) instead of square errors (NRMSE) and taking the median instead of the mean of the NRMSE (or NMAE) across scenarios did not substantially change the ranking of the models, still placing simple ARIMA models at the top of the list. However, while the choice of the mean strongly penalizes the most difficult scenarios, the median simply excludes these particularly complicated situations from the evaluation. For this reason, we have focused mainly on the mean NRMSE, as it allows us to favor models and methods that do not show aberrant behavior in case of trend reversal or stabilization, as this is becoming increasingly common (e.g., prostate, breast, and skin melanoma) and a similar pattern will hopefully be observed in the future for other cancer sites. Finally, looking at the prediction intervals did not alter our conclusions, with simple ARIMA models still being the best performing models.

Our first observation is that the models that offer the best fit to the data, for example, models selected using AIC, are not necessarily the best for prediction. The use of AIC may lead to the choice of overly complex models, for example, GLM or APC including complex interactions with splines, which are likely to overfit the data and are indeed less efficient for prediction than simpler and smoother models. The second observation is in line with the first one by advocating for simplicity. Approaches such as APC or BAPC estimate the effects of age, period, and/or cohort separately, considered as proxies for epidemiological risk factors, such as smoking and other carcinogens. However, when it is a matter of prediction, we have shown that better performance is obtained by simply extrapolating the trends using models such as ARIMA (or even simple regression models). As Booth and Tickle (2008) have already noted in the context of mortality projection, an extrapolation-based approach is often preferable to an explanation- or interpretation-based approach.

Our study has some limitations, all of which could motivate future work. For example, although we considered many models, we did not include overdispersion in our Poisson models. While this would affect the coverage of the prediction intervals, it would not greatly change the predictions and thus the NRMSE, our main criterion. More generally, many other prediction models and options can be formulated, for example, using quasi-Poisson or negative binomial distributions or trying other link functions, and it could become an interesting challenge for researchers to try to identify one that can outperform ARIMA in comparable settings. Second, we did not take into account the effects of prevention/screening programs and changes in medical technology and practice on cancer incidence rates. These factors can strongly influence the (past and future) rates and help identify trend changes (Etzioni et al., 2013). Such effects are however challenging to analyze, as they evolve over time. For example, the start of a cancer screening program leads to a short-term initial increase in incidence, followed by a stabilization or sometimes a decrease compared to past trends. The magnitude of these temporal effects will depend on the adherence to these prevention programs. Considering the impact of screening when making predictions is a subject of ongoing work. Another limitation is that we did not account for data underreporting and its evolution over time, which may introduce artifacts into predictions. In fact, while a high degree of completeness has been evaluated in Swiss cancer registration (Lorez et al., 2017), a slight underreporting of some cancers cannot be discarded. Finally, we have only modeled the most common cancers. Although we assume that simpler methods, such as ARIMA, should be suitable for rare cancers, as they are less prone to overfitting, we have not studied the prediction performance for these cancers and cannot draw any conclusions on this point. A specific analysis of the most suitable prediction methods for rare cancers would be another important subject of future work. The same consideration can be made for the prediction of specific age incidence rates.

In conclusion, we recommend using lower order ARIMA models to predict cancer incidence, for example, the ARIMA (2,1,1) or (1,1,0) models, which achieved the best overall performance in our comparison study. We suggest not using AIC to select the model, as this appears not to be the best strategy for prediction because it often results in overfitting. While APC and BAPC models remain the best to help interpret changes in cancer incidence trends, we recommend avoiding the widely used APC model for prediction, especially when the extrapolation is restricted to the drift. Finally, given the large uncertainty, particularly in the case of a trend reversal, we do not recommend predicting cancer incidence for periods far into the future but rather updating predictions regularly.

ACKNOWLEDGMENT

This work has been supported by the Krebsliga Schweiz (KFS-4385-02-2018) (Isabella Locatelli).

Open access funding provided by Universite de Lausanne.

CONFLICT OF INTEREST STATEMENT

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

Population data are available on the FSO website (https://www.pxweb.bfs.admin.ch/pxweb/fr/px-x-0102020000_104/px-x-0102020000_104/px-x-0102020000_104.px) and cancer incidence data can be requested from cancer registries after ethical approval of a project. In the Supporting Information, a simulated dataset, close to the original one is available, as well as the R code used to produce the analysis, tables, and figures presented in the article.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article were reproduced partially due to data confidentiality issues.

ORCID

Bastien Trächsel https://orcid.org/0000-0002-4519-094X

Jean-Luc Bulliard https://orcid.org/0000-0001-9750-2709

REFERENCES

Booth, H., & Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, *3*(1–2), 3–43. https://doi.org/10.1017/S1748499500000440

Bürkner, P.-C., Gabry, J., & Vehtari, A. (2020). Approximate leave-future-out cross-validation for Bayesian time series models. *Journal of Statistical Computation and Simulation*, *90*(14), 2499–2523. https://doi.org/10.1080/00949655.2020.1783262

Cameron, A. C., & Trivedi, P. K. (2013a). 2.4.4 Generalized linear models. In *Regression analysis of count data* (2nd ed., pp. 35–38). Cambridge University Press. https://doi.org/10.1017/CBO9781139013567

Cameron, A. C., & Trivedi, P. K. (2013b). 3 Basic count regression. In *Regression Analysis of Count Data* (2nd ed., p. 566). Cambridge University Press. https://doi.org/10.1017/CBO9781139013567

Cameron, J. K., & Baade, P. (2021). Projections of the future burden of cancer in Australia using Bayesian age-period-cohort models. *Cancer Epidemiology*, *72*, 101935. https://doi.org/10.1016/j.canep.2021.101935

Carstensen, B. (2007). Age–period–cohort models for the Lexis diagram. *Statistics in Medicine*, *26*(15), 3018–3045. https://doi.org/10.1002/sim.2764

Carstensen, B., Plummer, M., Laara, E., & Hills, M. (2022). *Epi: A package for statistical analysis in epidemiology*. https://CRAN.R-project.org/package=Epi

Chen, H. S., Portier, K., Ghosh, K., Naishadham, D., Kim, H.-J., Zhu, L., Pickle, L. W., Krapcho, M., Scoppa, S., Jemal, A., & Feuer, E. J. (2012). Predicting US and state-level cancer counts for the current calendar year: Part I: Evaluation of temporal projection methods for mortality. *Cancer*, *118*(4), 1091–1099. https://doi.org/10.1002/cncr.27404

Chen, W.-Q., Zheng, R.-S., & Zeng, H.-M. (2011). Bayesian age-period-cohort prediction of lung cancer incidence in China: Predicted lung cancer incidence in China. *Thoracic Cancer*, *2*(4), 149–155. https://doi.org/10.1111/j.1759-7714.2011.00062.x

Clayton, D., & Schifflers, E. (1987). Models for temporal variation in cancer rates. II: Age–period–cohort models. *Statistics in Medicine*, *6*(4), 469–481. https://doi.org/10.1002/sim.4780060406

Clements, M. S. (2005). Lung cancer rate predictions using generalized additive models. *Biostatistics*, *6*(4), 576–589. https://doi.org/10.1093/biostatistics/kxi028

Coupland, V. H., Okello, C., Davies, E. A., Bray, F., & Moller, H. (2010). The future burden of cancer in London compared with England. *Journal of Public Health*, *32*(1), 83–89. https://doi.org/10.1093/pubmed/fdp082

Dunn, P. K., & Smyth, G. K. (2018). Chapter 12: Tweedie GLMs. In P. K. Dunn & G. K. Smyth, *Generalized linear models with examples in R* (pp. 457–490). Springer. https://doi.org/10.1007/978-1-4419-0118-7_12

Earnest, A., Evans, S. M., Sampurno, F., & Millar, J. (2019). Forecasting annual incidence and mortality rate for prostate cancer in Australia until 2022 using autoregressive integrated moving average (ARIMA) models. *BMJ Open*, *9*(8), e031331. https://doi.org/10.1136/bmjopen-2019-031331

Etzioni, R., Gulati, R., Mallinger, L., & Mandelblatt, J. (2013). Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. *Annals of Internal Medicine*, *158*(11), 831. https://doi.org/10.7326/0003-4819-158-11-201306040-00008

Fritz, A., Percy, C., Jack, A., Shanmugaratnam, K., Sobin, L. H., Parkin, D. M., Whelan, S. L., & World Health Organization. (2000). *International classification of diseases for oncology*. [Classification internationale des maladies pour l'oncologie]. WHO IRIS. https://apps.who.int/iris/handle/10665/42344

Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*(2), 243–268. https://doi.org/10.1111/j.1467-9868.2007.00587.x

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691042893/time-series-analysis

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer. https://doi.org/10.1007/978-0-387-84858-7

Holford, T. R. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics*, *39*(2), 311–324. https://doi.org/10.2307/2531004

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, *87*(419), 659–671. https://doi.org/10.1080/01621459.1992.10475265

Lerman, P. M. (1980). Fitting segmented regression models by grid search. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *29*(1), 77–84. https://doi.org/10.2307/2346413

Li, C., Zhu, Y., Yang, J., Xu, D., Wang, J., Chen, K., & Li, Q. (2022). Incidence of lung cancer in Jiashan, Zhejiang province: Trend analysis from 1987 to 2016 and projection from 2017 to 2019. *Journal of Zhejiang University (Medical Sciences)*, *47*(4), 367–373. https://doi.org/10.3785/j.issn.1008-9292.2018.08.07

Lin, H., Shi, L., Zhang, J., Zhang, J., & Zhang, C. (2021). Epidemiological characteristics and forecasting incidence for patients with breast cancer in Shantou, Southern China: 2006–2017. *Cancer Medicine*, *10*(8), 2904–2913. https://doi.org/10.1002/cam4.3843

Lindgren, F., & Rue, H. (2008). On the second-order random walk model for irregular locations: RW2 for irregular locations. *Scandinavian Journal of Statistics*, *35*(4), 691–700. https://doi.org/10.1111/j.1467-9469.2008.00610.x

Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, *63*(19), 1–25. https://doi.org/10.18637/jss.v063.i19

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(4), 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Lorez, M., Bordoni, A., Bouchardy, C., Bulliard, J.-L., Camey, B., Dehler, S., Frick, H., Konzelmann, I., Maspoli, M., Mousavi, S. M., Rohrmann, S., & Arndt, V. (2017). Evaluation of completeness of case ascertainment in Swiss cancer registration. *European Journal of Cancer Prevention*, *26*, S139–S146. https://doi.org/10.1097/CEJ.0000000000000380

Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. (2013). Bayesian computing with INLA: New features. *Computational Statistics & Data Analysis*, *67*, 68–83. https://doi.org/10.1016/j.csda.2013.04.014

Møller, B., Fekjaer, H., Hakulinen, T., Sigvaldason, H., Storm, H. H., Talbäck, M., & Haldorsen, T. (2003). Prediction of cancer incidence in the Nordic countries: Empirical comparison of different approaches: Comparison of methods for incidence prediction. *Statistics in Medicine*, *22*(17), 2751–2766. https://doi.org/10.1002/sim.1481

Møller, B., Fekjaer, H., Hakulinen, T., Tryggvadóttir, L., Storm, H. H., Talbäck, M., & Haldorsen, T. (2002). Prediction of cancer incidence in the Nordic countries up to the year 2020. *European Journal of Cancer Prevention*, *11*(Suppl 1), S1–96.

Møller, B., Weedon-Fekjær, H., & Haldorsen, T. (2005). Empirical evaluation of prediction intervals for cancer incidence. *BMC Medical Research Methodology*, *5*(1), 21. https://doi.org/10.1186/1471-2288-5-21

NSW Cancer Institute. (2016). *Cancer incidence and projections 2011–2021*. Cancer Institute.

OFS. (2020). *Cancer, nouveaux cas et décès: Nombre, taux et évolution par localisation cancéreuse et période - 1987–2017 | Tableau*. Office fédéral de la statistique. https://www.bfs.admin.ch/asset/fr/14816247

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rahib, L., Wehner, M. R., Matrisian, L. M., & Nead, K. T. (2021). Estimated projection of US cancer incidence and death to 2040. *JAMA Network Open*, *4*(4), e214708. https://doi.org/10.1001/jamanetworkopen.2021.4708

Rapiti, E., Guarnori, S., Pastoors, B., Miralbell, R., & Usel, M. (2014). Planning for the future: Cancer incidence projections in Switzerland up to 2019. *BMC Public Health*, *14*(1), 102. https://doi.org/10.1186/1471-2458-14-102

Riebler, A., & Held, L. (2017). Projecting the future burden of cancer: Bayesian age-period-cohort analysis with integrated nested Laplace approximations: Projecting the future burden of cancer. *Biometrical Journal*, *59*(3), 531–549. https://doi.org/10.1002/bimj.201500263

Riebler, A., Held, L., & Rue, H. (2012). Estimation and extrapolation of time trends in registry data—Borrowing strength from related populations. *The Annals of Applied Statistics*, *6*(1), 304–333. https://doi.org/10.1214/11-AOAS498

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511755453

Rutherford, M. J., Thompson, J. R., & Lambert, P. C. (2012). Projecting cancer incidence using age-period-cohort models incorporating restricted cubic splines. *The International Journal of Biostatistics*, *8*(1),33. https://doi.org/10.1515/1557-4679.1411

Schmid, V., & Held, L. (2004). Bayesian extrapolation of space-time trends in cancer registry data. *Biometrics*, *60*(4), 1034–1042. JSTOR. https://doi.org/10.1111/j.0006-341X.2004.00259.x

Schumacher, D., & Jombart, T. (2021). *trending: Model temporal trends.* https://CRAN.R-project.org/package=trending

Shi, J., Cao, M., Wang, Y., Bai, F., Lei, L., Peng, J., Feletto, E., Canfell, K., Qu, C., & Chen, W. (2021). Is it possible to halve the incidence of liver cancer in China by 2050? *International Journal of Cancer*, *148*(5), 1051–1065. https://doi.org/10.1002/ijc.33313

Spiegelman, M., & Marks, H. H. (1966). Empirical testing of standards for the age adjustment of death rates by the direct method. *Human Biology*, *38*(3), 279–292.

Tsoi, K. K. F., Hirai, H. W., Chan, F. C. H., Griffiths, S., & Sung, J. J. Y. (2017). Cancer burden with ageing population in urban regions in China: Projection on cancer registry data from World Health Organization. *British Medical Bulletin*, *121*(1), 83–94. https://doi.org/10.1093/bmb/ldw050

Wong, M. C. S., Huang, J., Lok, V., Wang, J., Fung, F., Ding, H., & Zheng, Z.-J. (2021). Differences in incidence and mortality trends of colorectal cancer worldwide based on sex, age, and anatomic location. *Clinical Gastroenterology and Hepatology*, *19*(5), 955–966. e61. https://doi.org/10.1016/j.cgh.2020.02.026

Yang, L., Parkin, D. M., Ferlay, J., Li, L., & Chen, Y. (2005). Estimates of cancer incidence in China for 2000 and projections for 2005. *Cancer Epidemiology, Biomarkers & Prevention*, *14*(1), 243–250. https://doi.org/10.1158/1055-9965.243.14.1

Yu, X. Q., Luo, Q., Hughes, S., Wade, S., Caruana, M., Canfell, K., & O'Connell, D. L. (2019). Statistical projection methods for lung cancer incidence and mortality: A systematic review. *BMJ Open*, *9*(8), e028497. https://doi.org/10.1136/bmjopen-2018-028497

Zemni, I., Kacem, M., Dhouib, W., Bennasrallah, C., Hadhri, R., Abroug, H., Ben Fredj, M., Mokni, M., Bouanene, I., & Belguith, A. S. (2022). Breast cancer incidence and predictions (Monastir, Tunisia: 2002–2030): A registry-based study. *PLoS ONE*, *17*(5), e0268035. https://doi.org/10.1371/journal.pone.0268035

Zheng, Y., Zhang, L., Zhu, X., & Guo, G. (2020). A comparative study of two methods to predict the incidence of hepatitis B in Guangxi, China. *PLoS ONE*, *15*(6), e0234660. https://doi.org/10.1371/journal.pone.0234660

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Trächsel, B., Rousson, V., Bulliard, J.-L., & Locatelli, I. (2023). Comparison of statistical models to predict age-standardized cancer incidence in Switzerland. *Biometrical Journal*, e2200046. https://doi.org/10.1002/bimj.202200046

PLOS GLOBAL PUBLIC HEALTH

RESEARCH ARTICLE

Predicting the burden of cancer in Switzerland up to 2025

Bastien Trächsel[1]*, Elisabetta Rapiti[2], Anita Feller[3,4], Valentin Rousson[1], Isabella Locatelli[1☯], Jean-Luc Bulliard[1☯]

1 Center for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne, Switzerland, **2** Geneva Cancer Registry, University of Geneva, Geneva, Switzerland, **3** Foundation National Institute for Cancer Epidemiology and Registration (NICER), Zurich, Switzerland, **4** National Agency for Cancer Registration (NACR) Operated by NICER, Zurich, Switzerland

☯ These authors contributed equally to this work.
* sui.bastien@gmail.com

Citation: Trächsel B, Rapiti E, Feller A, Rousson V, Locatelli I, Bulliard J-L (2022) Predicting the burden of cancer in Switzerland up to 2025. PLOS Glob Public Health 2(10): e0001112. https://doi.org/10.1371/journal.pgph.0001112

Editor: Nnodimele Onuigbo Atulomah, Babcock University, NIGERIA

Received: July 11, 2022

Accepted: September 20, 2022

Published: October 14, 2022

Data Availability Statement: The Swiss cancer data used in these analyses were supplied by National Agency for Cancer Registration (NACR) and its partner registries in cantons Aargau, Basel, Bern, Fribourg, Geneva, Glarus & Graubünden, Luzern, Neuchâtel & Jura, St. Gallen & Appenzell, Thurgau, Ticino, Valais, Vaud, Zürich & Zug. Data access can be requested from NACR. Contact information for the data access request can be found at https://www.nacr.ch/en/data/ The contact email is requests[at]nicer.org.

Abstract

Predicting the short-term evolution of the number of cancers is essential for planning investments and allocating health resources. The objective of this study was to predict the numbers of cancer cases and of the 12 most frequent cancer sites, and their age-standardized incidence rates, for the years 2019–2025 in Switzerland. Projections of the number of malignant cancer cases were obtained by combining data from two sources: forecasts of national age-standardized cancer incidence rates and population projections from the Swiss Federal Statistical Office. Age-standardized cancer incidence rates, approximating the individual cancer risk, were predicted by a low-order Autoregressive Integrated Moving Average (ARIMA) model. The contributions of changes in cancer risk (epidemiological component) and population aging and growth (demographic components) to the projected number of new cancer cases were each quantified. Between 2018 and 2025, age-standardized cancer incidence rates are predicted to stabilize for men and women at around 426 and 328/100,000, respectively (<1% change). These projected trends are expected for most cancer sites. The annual number of cancers is expected to increase from 45,676 to 52,552 (+15%), more so for men (+18%) than for women (+11%). These increases are almost entirely due to projected changes in population age structure (+12% for men and +6% for women) and population growth (+6% for both sexes). The rise in numbers of expected cancers for each site is forecast to range from 4.15% (thyroid in men) to 26% (bladder in men). While ranking of the three most frequent cancers will remain unchanged for men (1st prostate, 2nd lung, 3rd colon-rectum), colorectal cancer will overtake by 2025 lung cancer as the second most common female cancer in Switzerland, behind breast cancer. Effective and sustained prevention measures, as well as infrastructural interventions, are required to counter the increase in cancer cases and prevent any potential shortage of professionals in cancer care delivery.

Funding: This work was supported by the Swiss Cancer Research Foundation (KFS-4385-02-2018) (IL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Predicting the future burden of cancer in a country has a pivotal role in cancer control planning. It provides evidence for policymakers seeking to allocate resources (e.g., cancer care, prevention and research) and to assess the economic burden of cancer. The future number of new cancer cases will depend on three main components. The first is an epidemiological component, i.e. the change in age-standardized cancer incidence rates, which approximates the change in individual risk factors as well as in preventive actions taken to counter them [1]. The other two components are demographic, i.e. the growth and aging of the population observed in many countries.

In general, prediction of the future number of cancer cases in a country is made by combining a prediction of age-standardized cancer incidence rates, obtained by models that attempt to identify and extend past trends into the future, with demographic projections of population size and age structure [2]. Methods to predict age-standardized cancer incidence rates most often used Age-Period-Cohort (APC) models, considering these three dimensions as proxies for different risks and preventive factors [2–5]. More recently, methods such as Bayesian Age-Period-Cohort (BAPC) models [6, 7], joinpoint regression [8, 9] and machine learning [10] have also been used. The choice of the statistical method can, however, have a significant impact on prediction. A recent study comparing a large set of prediction models showed the superiority of ARIMA (Auto Regressive Integrated Moving Average) methods which consistently outperformed more sophisticated methods [11].

In Switzerland, the National Agency for Cancer Registration (NACR) gathers data from regional cancer registries and produces national incidence figures, extrapolating the expected number of cancers for the few areas not covered by population-based registries [12, 13]. A study used these data for the period 1989–2009 to predict, by APC modelling, the evolution of age-standardized incidence rates up to the period 2015–19 for all cancers combined and for each of the 12 most frequent cancer sites [14]. Predicted age-standardized cancer incidence rates were then combined with population projections from the Swiss Federal Statistical Office (FSO) to forecast the number of new cancer cases to 2015–19. Eight years later, with Swiss cancer incidence series available up to the year 2018, these predictions can be compared with the observed trends, while new predictions are needed.

The first objective of the present study was to predict the age-standardized cancer incidence rates and the absolute number of new cancer cases in Switzerland, for all cancers combined and for each of the 12 most frequent cancers from 2019 until 2025. This was done combining NACR incidence data and Swiss population demographics projections. For the statistical analysis, we used a low order Auto Regressive Integrated Moving Average (ARIMA) model [15] for predicting age-standardized cancer incidence rates, as recommended by [11], while adopting the FSO projections for the demographic evolution. A second objective was to quantify the contribution of the change in the individual cancer risk, as approximated by age-standardized cancer incidence rates, and of population aging and growth to the expected change in the number of new cancer cases between 2018 and 2025.

Method

Cancer incidence data

As primary source of data, we used cancer incidence data that were calculated and produced by the Swiss National Agency for Cancer Registration (NACR) for all Switzerland. For this, NACR gathered data recorded by regional population-based cancer registries in Switzerland and extrapolated national figures including regions not covered by a registry. NACR made

available for years y = 1987–2018 the age standardized cancer rates λ_y^s at the European standard population *s* [16]:

$$\lambda_y^s = \sum_{a \in A} \lambda_{y,a} P_{s,a} / P_s \quad (1)$$

In (1), *A* is a partition of the ages into 5-year age groups (i.e. 0–4, 5–9, . . ., last open group 85 years and over), $\lambda_{y,a}$ is the age-specific incidence rate for the year y and age group *a* ($a \in A$) (e.g. *a* = 0–4 years), and $P_{s,a}/P_s$ is the proportion of people in the age group *a* in the standard population *s*. We considered age-standardized cancer incidence rates for the 12 most frequent cancer sites according to the ICD10 codes: oral cavity and pharynx (C00-14), stomach (C16), colon and rectum (C18-20), lung-bronchus-trachea (C33-34), skin melanoma (C43), breast (C50), corpus uteri and uterus NOS (C54-55), ovary (C56), prostate (C61), bladder (C67), thyroid (C73), non-Hodgkin lymphoma (C82-85, C96). All other cancers were regrouped in the category "other" (All other ICD10 codes except C44) to allow calculation of the overall cancer incidence (C00-43, C45-97).

Demographic data

Our projections of cancer incidence based on NACR data were combined with demographic projections for the Swiss population. For the evolution of the Swiss population from 2021 up to 2025, we used the so-called "average A00-2020" reference scenario of the Swiss Federal Statistical Office (FSO) [17]. It predicts globally and by age groups the most plausible growth of the population based on demographic changes in mortality, migration, and fertility. This reference scenario predicts a number of women increasing from 4,292,551 in 2018 to 4,538,813 in 2025 (+5.74%) and a number of men increasing from 4,221,778 in 2018 to 4,482,174 in 2025 (+6.17%). For years up to 2020, we used the official population numbers from the FSO.

Modelling and prediction of age-standardized cancer incidence rates

For the choice of a statistical method to predict age-standardized cancer incidence rates, we referred to a recent comparison of the accuracy of several models and methods, based on the repeated application of leave-future-out cross-validation on a 35-year series of incidence data from cancer registries [11]. According to [11], the best predictions are obtained by simply extrapolating trends in standardized cancer rates (1) by ARIMA (Auto Regressive Integrated Moving Average, [15]) models. These models showed better accuracy than the more complex and widely used APC methods and have the advantage of requiring only the age-standardized rates, while the age-specific incidence rates are not needed. The simple ARIMA(2,1,1) proved to be the best performing method and was adopted for the present study. Mathematically, this model can be described as follows. ARIMA(2,1,1) combines a differenced 2-order autoregressive model with a 1-order moving average model, which can be written as:

$$d\lambda_y^s = c + \alpha_1 d\lambda_{y-1}^s + \alpha_2 d\lambda_{y-2}^s + \varepsilon_y + \theta_1 \varepsilon_{y-1} \quad (2)$$

In (2), $d\lambda_y^s$ (y = 1987–2018) are differences between age-standardized rates at two consecutive years, $d\lambda_y^s = \lambda_y^s - \lambda_{y-1}^s$, $\in_y$ are normally distributed random errors, and *c*, α_1, α_2, θ_1 are a constant, two auto-regressive coefficients and one moving average coefficient, respectively. This model has been fitted to the NACR data described above. Coefficients in model (2) were estimated separately for each cancer site and for each sex, when applicable. We used the ARIMA function from software R [18] to estimate these coefficients, to estimate (with 95% prediction intervals) age-standardized cancer incidence rates $\hat{\lambda}_y^s$ for years y = 1987–2018, and to predict

them (with 95% prediction intervals) for years y = 2019–2025. Technically, the estimation method implemented in R is based on [19].

The percent change in age-standardized cancer incidence rates between 2018 and 2025 was then estimated by comparing the predicted value for 2025, $\hat{\lambda}^s_{2025}$, with estimated value for 2018, $\hat{\lambda}^s_{2018}$, instead of the observed value λ^s_{2018}. This choice was intended to reduce the impact of random fluctuations in 2018, $\in_{2018}$, on our comparisons.

Prediction of the future number of new cancer cases

In order to predict the absolute number of new cases expected each year $\hat{N}_y$, y = 2019–2025, the predicted age-standardized rates $\hat{\lambda}^s_y$ should be conveniently combined with the FSO population forecasts by 5-year age classes for the same years. We used the following approximation:

$$\hat{N}_y \approx P_y \cdot \hat{\lambda}^s_y \cdot \lambda^y_{y^\star} / \lambda^s_{y^\star} \quad (3)$$

Here, y^* is any year in the period 1987–2018 (we considered $y^\star = 2018$); y is a future year for which we wish to make the prediction (y = 2019–2025); P_y is the total population for year y according to the FSO forecast [17]; $\hat{\lambda}^s_y$ are predicted standardized rates for year y according to model (2); $\lambda^y_{y^\star} = \sum_{a\in A} \lambda_{y^\star,a} P_{y,a}/P_y$ (NACR) are cancer rates of year $y^\star$ standardized on the population structure at future year y (FSO forecast); and $\lambda^s_{y^\star}$ are cancer rates in year $y^\star$ standardized on the European standard population s. Approximation (3) is valid provided that the rate ratio between any two age groups $a_1, a_2 \in A$ remains approximately stable over time, that is: $\lambda_{y,a_1}/\lambda_{y,a_2} \approx \lambda_{y^\star,a_1}/\lambda_{y^\star,a_2}$; $y \in 2019–2025$, $y^* = 2018$. In that case, the ratio of standardized rates for years y and y^* will not be much affected by the particular choice of standardization, so that $\hat{\lambda}^s_y/\lambda^s_{y^\star} \approx \hat{\lambda}^y_y/\lambda^y_{y^*}$, ensuring validity of (3). Empirically this is generally the case, as shown by [20].

Factorization of the predicted change in number of new cancer cases

The predicted change in the number of new cancer cases between 2018 and 2025 was factorized into three components [21]. Considering that $N_{2018} = P_{2018}\lambda^{2018}_{2018}$, the change (in %) of the number of new cancer cases between 2018 and 2025 can be expressed as:

$$D = 100 \cdot \left[\hat{N}_{2025}/N_{2018} - 1\right]\% = 100 \cdot [(1 + A/100) \cdot (1 + B/100) \cdot (1 + C/100) - 1]\% \quad (4)$$

In (4), $A = 100(\hat{\lambda}^s_{2025}/\lambda^s_{2018} - 1)$ is the change (in %) in the number of new cancer cases due to the change in individual cancer risk, as approximated by age-standardized cancer rates, $B = 100(\lambda^{2025}_{2018}/\lambda^{2018}_{2018} - 1)$ is the change (in %) in the number of new cancer cases due to the population ageing, and $C = 100(P_{2025}/P_{2018} - 1)$ the change (in %) due to the population growth. This factorization allows us to quantify the multiplicative contributions of the epidemiological component (A) and the two demographic components (B and C) on the expected future number of new cancer cases (D).

Results

Over the past decade, we have observed in Switzerland a stabilization of age-standardized cancer incidence rates and a steady increase in the number of new cancer cases, as previously predicted [14]. Our new predictions for age-standardized cancer incidence rates and absolute number of cancer cases up to 2025 point to a continuation of these trends (Figs 1 and 2).

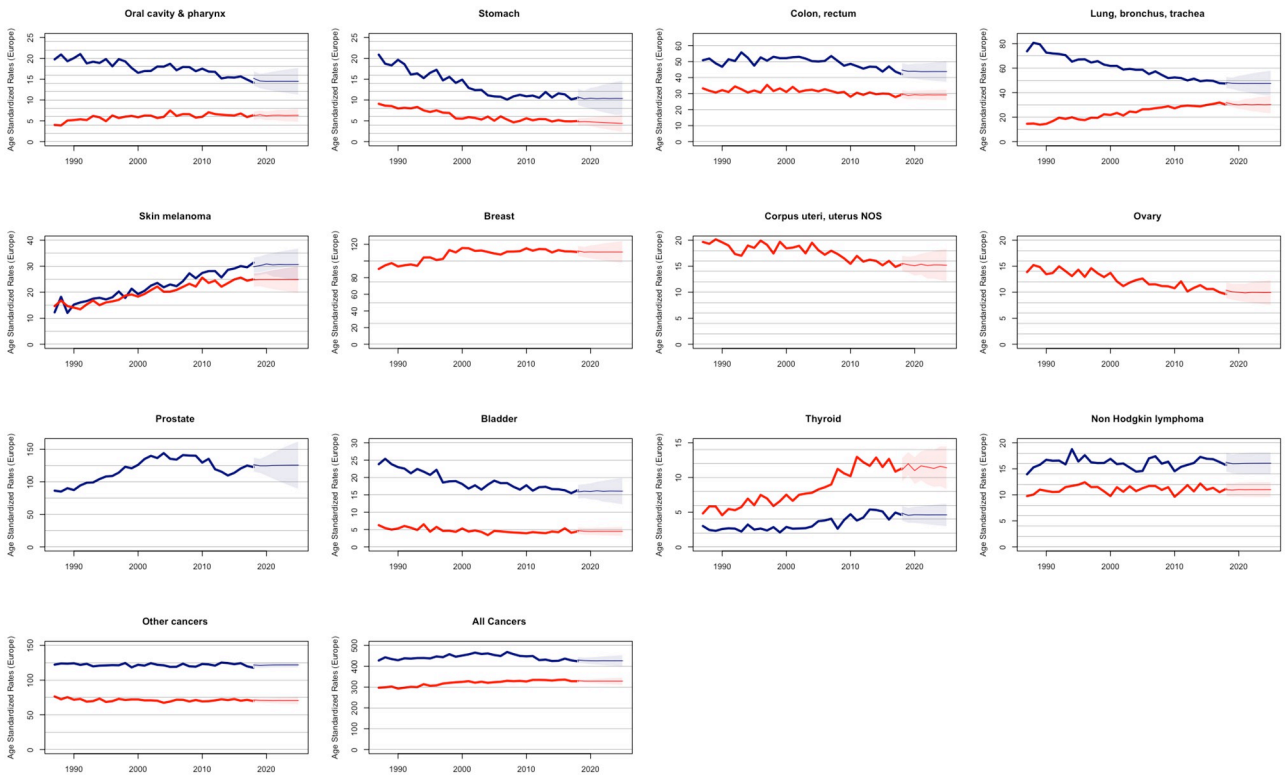

Fig 1. Observed and projected age standardized rates for all cancers and 12 cancer sites in Switzerland. Bold lines represent actual rates; thinner lines represent projected rates; shaded areas represent 95% prediction intervals. Men in blue; women in red.

https://doi.org/10.1371/journal.pgph.0001112.g001

While age-standardized rates are expected to remain roughly stable in the coming years (Fig 1), the number of cancer cases is predicted to increase substantially for each cancer site and for all cancers combined (Fig 2).

Table 1 shows the expected change in the estimated number of new cancer cases between 2018 and 2025 for each cancer site and sex. This change is also factorized into its epidemiological and demographic components, as explained in the Method section, Eq (4). The age-standardized cancer incidence rates, reflecting changes in cancer risk, are overall expected to remain roughly stable until 2025 for both sexes (see also Fig 1). For men, a slight increase in incidence rate is expected for cutaneous melanoma (+2.75%), and bladder cancer (+1.24%), whereas the incidence rates are expected to decrease slightly for cancers of the lung, bronchus and trachea (-1.38%), oral cavity and pharynx (-4.61%), stomach (-3.72%), colon-rectum (-2.15%), thyroid (-5.07%), and for non-Hodgkin's lymphoma (-1.22%) and to remain approximately constant for prostatic cancer (-0.66%). For women, a slight increase in standardized incidence rate is foreseen for cancers of the oral cavity and pharynx (+1.99%) and thyroid (+2.91%), while a slight decrease is predicted for cancers of the colon-rectum (-1.98%), lung, bronchus and trachea (-4.86%), breast (-1.08%), uterus (-2.24%), ovary (-4.17%) and bladder (-5.09%), and more markedly for stomach cancer (-8.13%). No change in incidence rate is expected for cutaneous melanoma (+0.68%) and non-Hodgkin's lymphoma (+0.17%) in Swiss women.

The total number of new cancer cases is expected to increase between 2018 and 2025 from 24,987 to 29,553 (+18%) for men and from 20,689 to 22,999 (+11%) for women, i.e. an absolute increase of 6,876 new cancer cases (+15%) for both sexes. This increase is almost entirely

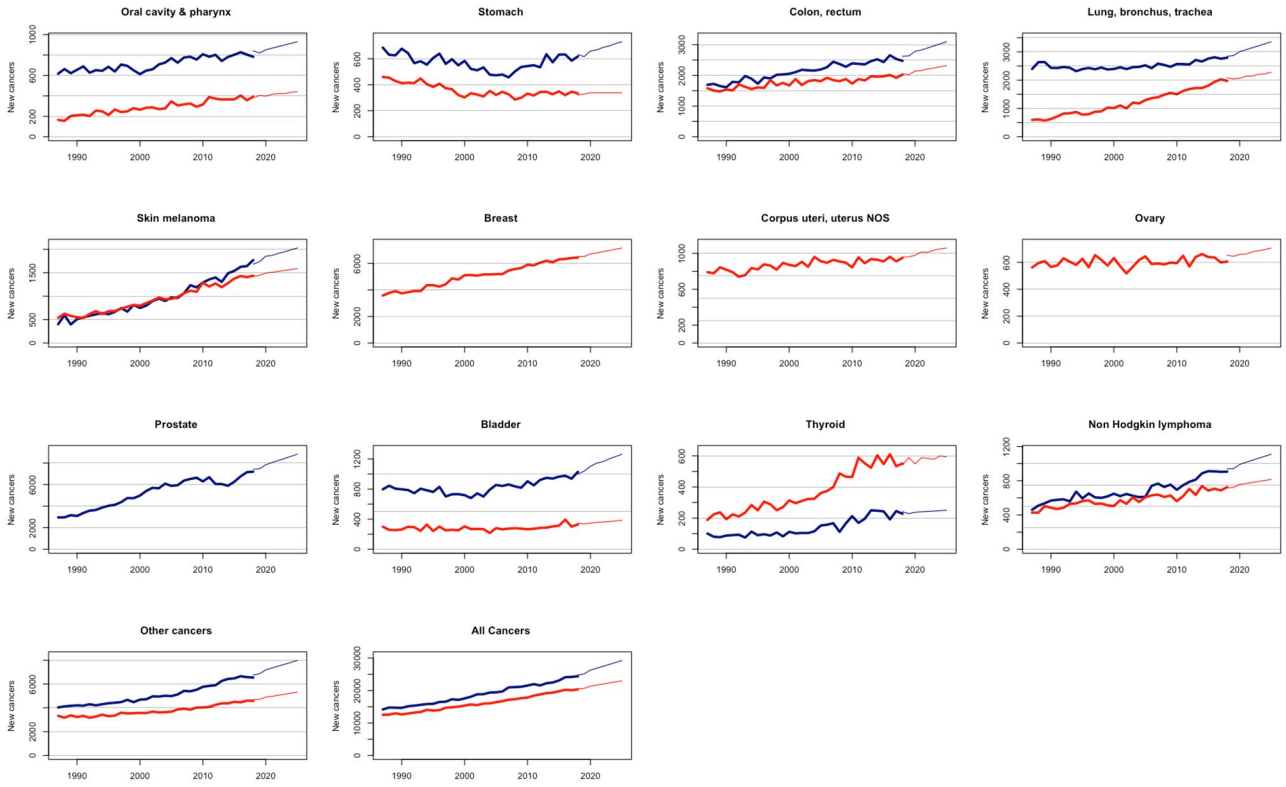

Fig 2. Observed and projected numbers of new cancer cases for all cancers and 12 cancer sites in Switzerland. Bold lines represent actual numbers; thinner lines represent projected numbers. Men in blue; women in red.

https://doi.org/10.1371/journal.pgph.0001112.g002

attributable to the projected aging (+12% for men and +6% for women) and growth (+6% for both sexes) of the Swiss population, since the contribution of the evolution of cancer risk amounts to less than 1% for both sexes. A rise in the number of new cancer cases is predicted for each cancer site and sex, with increases ranging from 4.15% for thyroid in men to 26% for bladder in men. Here also, all increases are primarily due to the demographic evolution, with the contribution of population aging generally outweighing that of population growth, with the notable exception of thyroid cancer for both sexes (Table 1). Only for thyroid cancer in women and cutaneous melanoma in men, an increase in risk partly contributed to the overall increase in the expected number of new cancer cases. Of note, for the few sites where we forecast a decrease in cancer risk (for example stomach cancer in women), this decrease is insufficient to offset the increase due to population growth and aging.

Discussion

In this study, we presented projections of the burden of cancer in Switzerland until 2025, in terms of both age-standardized rates and number of new cases, using, for the first time, the simple ARIMA method. Altogether, cancer incidence is predicted to stabilize for men and women, and for most cancer sites. The number of cases is, however, expected to increase by 15% (from 45,676 in 2018 to 52,552 in 2025), more so in absolute and relative terms for men than women. This increase is almost entirely driven by demographic changes. For men, it is primarily due to the projected population aging (+12.4%) followed by population growth (+6.2%) whereas the contributions of population aging and growth to the increase were comparable for women (6.0% and 5.7%, respectively).

Table 1. Estimated numbers of cancer cases for 2018 and predictions for 2025 in Switzerland by cancer site and by sex. Overall predicted changes are factorized (multiplicatively) into changes in standardized risk (A), in population structure (B), and in population size (C) (Eq (4)).

	MEN						WOMEN					
CANCER SITE	Number of cases 2018	Number of cases 2025	Change in risk A (%)	Change in pop structure B (%)	Change in pop size C (%)	Overall change (%)	Number of cases 2018	Number of cases 2025	Change in risk A (%)	Change in pop structure B (%)	Change in pop size C (%)	Overall change (%)
ORAL CAVITY & PHARYNX	841	929	-4.61	9.07	6.17	18.46	382	442	1.99	7.29	5.74	15.71
STOMACH	632	731	-3.72	13.15	6.17	15.66	321	338	-8.13	8.4	5.74	5.3
COLON, RECTUM	2625	3095	-2.15	13.49	6.17	17.9	2050	2319	-1.98	9.14	5.74	13.12
LUNG, BRONCHUS, TRACHEA	2844	3358	-1.38	12.76	6.17	18.07	2088	2280	-4.86	8.55	5.74	9.2
SKIN MELANOMA	1684	2029	2.75	10.45	6.17	20.49	1421	1589	0.68	5.04	5.74	11.82
BREAST							6515	7156	-1.08	5.01	5.74	9.84
CORPUS UTERI, UTERUS NOS							961	1060	-2.24	6.7	5.74	10.3
OVARY							654	706	-4.17	6.53	5.74	7.95
PROSTATE	7420	8815	-0.66	12.64	6.17	18.8						
BLADDER	1000	1262	1.24	17.41	6.17	26.2	346	383	-5.09	10.3	5.74	10.69
THYROID	241	251	-5.07	3.34	6.17	4.15	540	593	2.91	0.91	5.74	9.81
NON HODGKIN LYMPHOMA	937	1109	-1.22	12.85	6.17	18.36	716	815	0.17	7.47	5.74	13.83
OTHER CANCERS	6763	7974	0.07	10.98	6.17	17.91	4695	5318	-1.31	8.54	5.74	13.27
ALL CANCERS	24987	29553	-0.89	12.40	6.17	18.27	20689	22999	-0.78	5.96	5.74	11.17

https://doi.org/10.1371/journal.pgph.0001112.t001

For all cancers combined, the previous predictions until 2015–19 [14] closely corresponded to the observed Swiss figures until 2018 [22], i.e. a stable incidence rate for women and a slightly decreasing incidence rate for men, with an increasing number of cancers for both sexes. For a few sites such as stomach cancer in men and breast cancer in women, we observed, however, a less marked dynamic than was expected. This means a lesser decrease in incidence rate and therefore a larger increase in the number of cancers than predicted by [14]. For breast cancer, the implementation of mammography screening programs in several Swiss cantons since 2010, leading to the earlier detection of cancers that would otherwise be diagnosed a few years later, may explain the difference between predicted and observed breast cancer incidence rates [23].

Using the ARIMA method recommended by a large comparison study [11], we forecast a continuation until 2025 of the recent stable trends in age-standardized incidence rates for all cancers combined (around 426/100,000 in men and 328/100,000 in women) and a stabilization of incidence for most cancer sites considered. A stabilized age-standardized incidence can result from the continuation of past recent trends (colon-rectum, breast, prostate, bladder and non-Hodgkin lymphoma), the leveling off of a downward trend (for men: oral cavity, pharynx, lung and stomach; for women: corpus uteri, uterus, ovary and thyroid) or the leveling off of an upward trend (cutaneous melanoma for both sexes and oral cavity, pharynx and lung for women). While no substantial increase in incidence is forecast for any cancer site, the incidence of stomach cancer in women is predicted to substantially decrease (-8.13%), continuing

the recently reported downward trend in Switzerland [22]. Similarly, the incidence of thyroid cancer in men, which recorded the greatest increase in incidence of all cancer sites in Swiss men between 2008–2012 and 2013–2017 [22], is predicted to materially decrease (-5.07%) up to 2025.

Interpretation of cancer projection is challenging. The latency period between preventive measures aiming at reducing exposure to specific risk factors and its resulting effect on cancer incidence–and its magnitude–are difficult to anticipate. Further, the effect of early detection activities is often difficult to account for. This is particularly true for Switzerland where initiation of organized cancer screening programs and early detection activities mostly occur at regional level, without uniform or simultaneous application at the national level, and changes in clinical and diagnostic practices are virtually unpredictable. However, predicted incidence trends for several cancers are nevertheless largely corroborated by the Swiss epidemiological context. For example, the expected decrease, or stabilization after a long period of decrease, for males in incidence of cancers largely attributable to tobacco (oral cavity & pharynx, lung, bladder) and the predicted stabilization of incidence for those tobacco-associated cancers in females, after a long period of increase, are in line with the reduction of smoking prevalence and smoking ban policy in public places, which occurred later in Switzerland than in most other countries. The plateauing incidence rate of prostate cancer, the most common cancer in Swiss men, after years of a steady increase followed by a decrease has largely been attributable to changes in PSA screening and clinical workup practices [24]. Analogously, the strong rise between 1998 and 2012 in incidence of thyroid cancer in Switzerland, more pronounced in women than in men, was shown to be limited to small papillary carcinoma and likely due to overdiagnosis [25, 26]. The predicted stabilization of incidence in women and slight decrease in men may reflect recent favourable changes in clinical and diagnostic practices, as for prostate cancer. The predicted attenuation of the rise in melanoma incidence, after years of steady increase, has also been observed in countries with longstanding and sustained prevention activities against skin cancer [27–29]

In terms of predicted number of future new cases by 2025, the ranking of the three most common cancers will remain unchanged in the coming years for men (1st prostate, 2nd lung, 3rd colon-rectum), but colorectal cancer is expected to overtake by 2025 lung cancer as the second most frequent female cancer in Switzerland, behind breast cancer.

Comparing our results with recent projection studies from other industrialized countries, we note that a similar stabilization of incidence rates has overall been expected in 2020 (using data from 2010–12) in a study applying machine learning algorithms to predict incidence rates of all cancers and of the four main cancer types (lung, breast, prostate and colon) in Europe [10]. A study applying a BAPC model recently predicted for Australia an annual increase of 3% in the number of cancer cases between 2016 and 2031 [7]. This annual increase corresponds to an increase of $100(1.03^7 - 1)\% = 23\%$ over a 7-year period, which is higher than the 15% increase obtained in Switzerland over the 7-year period 2018–2025. This difference may partly be explained by a more marked dynamic of the age-standardized incidence rates predicted for Australia, i.e. a slight decrease for colon, stomach, and lung cancer in men and an increase for melanoma and lung cancer in women, in contrast to the stabilization expected for Switzerland. However, direct comparisons between countries should consider differences in exposure to risk factors, preventive measures, screening and intensity of diagnostic procedures, even if the evolution of the demographic components appears similar in Australia and Switzerland, with a continuous growth of life expectancy [30]. In this respect, the situation differs from the United States, where life expectancy has been stagnating or even decreasing in the past ten years. As demographic factors strongly influence the future number of cancers, this could explain, at least in part, the recent prediction of an overall relatively small rise from

1,735,000 new cancer cases in 2020 to 1,881,000 in 2040 in the United States [9], contrasting with our projections for Switzerland and those for Europe [10] and Australia [7].

This study has strengths and inherent limitations. One strength is the adoption of the ARIMA method for prediction. In addition to its demonstrated validity in a model comparison study based on a large number of leave-future-out cross-validation scenarios [11], this method can be directly applicable to standardized rate series without requiring ad hoc methods to mitigate the predicted dynamics (dumping or forced stabilization), as is often the case for APC models [5, 14]. However, as the ARIMA models predict standardized rates but not age-specific rates, an approximation [20] was necessary to combine these predictions with the demographic forecasts to obtain predicted numbers of cancer cases. A second strength of this study is that it uses annual rates rather than 5-year average rates for prediction, as is the more usual practice [5, 14]. Our approach allows us to use more detailed information and to predict future trends more accurately. However, this choice also entails a limitation when we want to quantify the variation of the rate between an observed year (2018) and a future year (2025), regarding the random fluctuations that may occur in a specific year (the year 2018 may present a rate slightly higher or lower than the trend, precisely due to this random variability). To mitigate this effect, we decided to compare the prediction for 2025 to the rate estimated by the model for 2018, rather than to the observed rate. As ARIMA models are based on moving averages, this choice smoothed out random fluctuations.

The major limitation of this study is the implicit assumption that no disruptive or unpredictable event will interfere with the continuation of past trends. This assumption, inherent to any prediction of the future from the past, has been violated twice. First, the Swiss cancer registration system was profoundly reformed in 2020 when notifications of all oncological diseases became mandatory. While the Cancer Registration Act (CRA) enabled full coverage of the population, constraints on patients informed consent and rights to veto prior to any case registration should affect the future completeness of registration to an—as yet—unknown degree [31, 32]. The magnitude of this disruption appears very likely to exceed any inaccuracy in the number of registered cancers (exhaustivity generally exceeds 95% in Swiss registries for most cancer sites [12]) or any imprecision in the extrapolation method used by NACR for regions not covered by cancer registry between 1987 and 2018. In this respect, our projections will provide a valuable baseline to estimate the effect of the CRA. Second, the unexpected COVID-19 pandemic produced some disruption in demographics in the last two years due to an increased mortality, particularly in older population groups where cancer risk is highest. Whether the reference scenario of the Swiss Federal Office for Statistics for population projection we used still holds true is unknown. Delayed cancer diagnosis due to the pandemic and the lockdown could, to some extent, also temporarily reduce cancer incidence and differently for various cancer types [33]. No model applied to pre-pandemic data could anticipate this kind of evolution. For these reasons, we believe that any prediction should not go too far into the future. We have limited ourselves to 7 years. Another limitation is the time lag inherent to complete cancer registration data which meant that, as of 2022, 2018 was the latest available year for predicting the Swiss cancer burden until 2025. The final limitation, which our study shares with all studies combining a prediction of incidence rates with population projections, is the difficulty to obtain prediction intervals for the absolute numbers of cancer cases as the uncertainty associated with the prediction of population figures is not available. This could be the subject of future work.

Despite these limitations, we can reasonably conclude that our simple prediction method based on ARIMA model was able to project trends in cancer incidence rate and absolute numbers of cancers until 2025 which overall concurred with the epidemiological context and knowledge in Switzerland. These trends generally pointed towards a stabilization of

age-standardized incidence rates for most cancers, accompanied by a substantial increase in the numbers of all types of cancer, driven by demographic changes. To counter the unavoidable increase in cancer cases, more effective and sustained prevention measures targeting factors such as obesity, physical activity, healthy diet and tobacco use are necessary. Structural interventions should also be devised in order to prevent any potential shortage of professionals in cancer care delivery.

Author Contributions

Conceptualization: Bastien Trächsel, Elisabetta Rapiti, Anita Feller, Valentin Rousson, Isabella Locatelli, Jean-Luc Bulliard.

Data curation: Bastien Trächsel, Elisabetta Rapiti, Anita Feller, Jean-Luc Bulliard.

Formal analysis: Bastien Trächsel, Anita Feller, Valentin Rousson, Isabella Locatelli.

Funding acquisition: Elisabetta Rapiti, Valentin Rousson, Isabella Locatelli, Jean-Luc Bulliard.

Investigation: Bastien Trächsel, Elisabetta Rapiti, Isabella Locatelli, Jean-Luc Bulliard.

Methodology: Bastien Trächsel, Elisabetta Rapiti, Valentin Rousson, Isabella Locatelli, Jean-Luc Bulliard.

Project administration: Bastien Trächsel, Isabella Locatelli.

Resources: Bastien Trächsel, Elisabetta Rapiti, Valentin Rousson, Isabella Locatelli, Jean-Luc Bulliard.

Software: Bastien Trächsel.

Supervision: Elisabetta Rapiti, Valentin Rousson, Isabella Locatelli, Jean-Luc Bulliard.

Validation: Bastien Trächsel, Elisabetta Rapiti, Isabella Locatelli, Jean-Luc Bulliard.

Visualization: Bastien Trächsel, Elisabetta Rapiti, Valentin Rousson, Isabella Locatelli, Jean-Luc Bulliard.

Writing – original draft: Bastien Trächsel, Elisabetta Rapiti, Isabella Locatelli, Jean-Luc Bulliard.

Writing – review & editing: Bastien Trächsel, Elisabetta Rapiti, Anita Feller, Valentin Rousson, Isabella Locatelli, Jean-Luc Bulliard.

References

1. Parkin D. M. (2006). The global health burden of infection-associated cancers in the year 2002. *International Journal of Cancer*, 118, 3030–3044. https://doi.org/10.1002/ijc.21731 PMID: 16404738
2. Moller B and Fekjær Harald and Hakulinen Timo and Tryggvadóttir Laufey and Storm Hans H and Talback M, et al. (2002). Prediction of cancer incidence in the Nordic countries up to the year 2020. *European Journal of Cancer Prevention*, 11.
3. Holford T. R. (1983). The Estimation of Age, Period and Cohort Effects for Vital Rates. *Biometrics*, 39, 311–324. Available from http://www.jstor.org/stable/2531004 PMID: 6626659
4. Carstensen B. (2007). Age-period-cohort models for the Lexis diagram. *Statistics in Medicine*, 26, 3018–3045. https://doi.org/10.1002/sim.2764 PMID: 17177166
5. Møller H., Fairley L., Coupland V., Okello C., Green M., Forman D., et al. (2007). The future burden of cancer in England: incidence and numbers of new patients in 2020. *British Journal of Cancer*, 96, 1484–1488. https://doi.org/10.1038/sj.bjc.6603746 PMID: 17473821
6. Schmid V., & Held L. (2004). Bayesian Extrapolation of Space-Time Trends in Cancer Registry Data. *Biometrics*, 60, 1034–1042. Available from http://www.jstor.org/stable/3695483 https://doi.org/10.1111/j.0006-341X.2004.00259.x PMID: 15606424

7. Cameron J. K., & Baade P. (2021). Projections of the future burden of cancer in Australia using Bayesian age-period-cohort models. *Cancer Epidemiology*, 72, 101935. https://doi.org/10.1016/j.canep.2021.101935 PMID: 33838461

8. Gillis D., & Edwards B. P. (2019). The utility of joinpoint regression for estimating population parameters given changes in population structure. *Heliyon*, 5, e02515. https://doi.org/10.1016/j.heliyon.2019.e02515 PMID: 31768426

9. Rahib L., Wehner M. R., Matrisian L. M., & Nead K. T. (2021, April). Estimated Projection of US Cancer Incidence and Death to 2040. JAMA Network Open, 4, e214708–e214708. https://doi.org/10.1001/jamanetworkopen.2021.4708 PMID: 33825840

10. Sekeroglu B., & Tuncal K. (2021). Prediction of cancer incidence rates for the European continent using machine learning models. *Health Informatics Journal*, 27, 1460458220983878. https://doi.org/10.1177/1460458220983878 PMID: 33506703

11. Trächsel B., Rousson V., Bulliard J., Locatelli I. (2022). Comparison of Statistical Models to Predict Cancer Incidence in Switzerland. Manuscript submitted for publication.

12. Lorez M., Bordoni A., Bouchardy C., Bulliard J.-L., Camey B., Dehler S., et al. (2017). Evaluation of completeness of case ascertainment in Swiss cancer registration. *European Journal of Cancer Prevention*, 26. https://doi.org/10.1097/CEJ.0000000000000380 PMID: 28574868

13. Raymond L., Bulliard J. L., Levi F., Enderlin F., Méan A., Schüler G., et al. (1994). Estimating the incidence of cancers in Switzerland: 1983–1987. *European Journal of Cancer*, 30, 978–982. https://doi.org/10.1016/0959-8049(94)90127-9 PMID: 7946595

14. Rapiti E., Guarnori S., Pastoors B., Miralbell R., & Usel M. (2014). Planning for the future: cancer incidence projections in Switzerland up to 2019. BMC Public Health, 14, 102. https://doi.org/10.1186/1471-2458-14-102 PMID: 24484472

15. Hamilton J. D., & Press P. U. (1994). *Time Series Analysis*. Princeton University Press.

16. Waterhouse JAH C. (1976). *Cancer incidence in five continents*. (Vol. 3). (Powell J., Ed.) Lyon: IARC.

17. BFS. (2020, November). *Les scénarios de l'évolution de la population de la Suisse et des cantons 2020–2050*. Neuchâtel: Federal Statistical Office (BFS). https://dam-api.bfs.admin.ch/hub/api/dam/assets/14963222/master

18. R Core Team, "R: A Language and Environment for Statistical Computing," Vienna, 2020.

19. Gardner G, Harvey A. C. and Phillips G. D. A. (1980). Algorithm AS 154: An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering. Applied Statistics, 29, 311–322. https://doi.org/10.2307/2346910

20. Spiegelman M and Marks H, Empirical testing of standards for the age adjustment of death rates by the direct method, *Human Biology*, vol. 38, p. 279–292, 1966.

21. Miettinen Olli S. (1972, August). COMPONENTS OF THE CRUDE RISK RATIO. *American Journal of Epidemiology*, 96, 168–172. https://doi.org/10.1093/oxfordjournals.aje.a121443 PMID: 4261806

22. BFS. (2021, October). *Le cancer en Suisse, rapport 2021—Etat des lieux et évolutions*. Neuchâtel: Federal Statistical Office (BFS). https://dam-api.bfs.admin.ch/hub/api/dam/assets/19305697/master

23. Bulliard J.-L., Braendle K., Fracheboud J., & Zwahlen M. (2021). Breast cancer screening programmes in Switzerland, 2010–2018. Tech. rep., Unisanté. Available from https://www.swisscancerscreening.ch/fr/federation/publications

24. Jegerlehner S., Chiolero A., Aujesky D., Rodondi N., Germann S., Konzelmann I., on behalf of the NICER Working Group. (2019, February). Recent incidence and surgery trends for prostate cancer: Towards an attenuation of overdiagnosis and overtreatment? *PLOS ONE*, 14, 1–11. https://doi.org/10.1371/journal.pone.0210434 PMID: 30716740

25. Jegerlehner S., Bulliard J.-L., Aujesky D., Rodondi N., Germann S., Konzelmann I., et al. (2017, June). Overdiagnosis and overtreatment of thyroid cancer: A population-based temporal trend study. *PLOS ONE*, 12, 1–12. https://doi.org/10.1371/journal.pone.0179387 PMID: 28614405

26. Wirth S., Syleouni M.-E., Karavasiloglou N., Rinaldi S., Korol D., Wannera M., et al. (2021). Incidence and mortality trends of thyroid cancer from 1980 to 2016. *Swiss Med Wkly*. https://doi.org/10.4414/smw.2021.w30029 PMID: 34751539

27. Whiteman D. C., Green A. C., & Olsen C. M. (2016). The Growing Burden of Invasive Melanoma: Projections of Incidence Rates and Numbers of New Cases in Six Susceptible Populations through 2031. *Journal of Investigative Dermatology*, 136, 1161–1171. https://doi.org/10.1016/j.jid.2016.01.035 PMID: 26902923

28. Aitken J. F., Youlden D. R., Baade P. D., Soyer H. P., Green A. C., & Smithers B. M. (2018). Generational shift in melanoma incidence and mortality in Queensland, Australia, 1995–2014. *International Journal of Cancer*, 142, 1528–1535. https://doi.org/10.1002/ijc.31141 PMID: 29105744

29. Garbe C., Keim U., Gandini S., Amaral T., Katalinic A., Hollezcek B., et al. (2021). Epidemiology of cutaneous melanoma and keratinocyte cancer in white populations 1943–2036. *European Journal of Cancer*, 152, 18–25. https://doi.org/10.1016/j.ejca.2021.04.029 PMID: 34062483

30. Eurostat Union European, (2022, March). Population on 1 January by age group and sex. *Population on 1 January by age group and sex*. Available from https://ec.europa.eu/eurostat/databrowser/view/demo_pjangroup/default/table?lang=en

31. Martani A., Erard F., Casonato C., & Elger B. S. (2021). The devil is in the details: an analysis of patient rights in Swiss cancer registries. *Journal of Medical Ethics*. https://doi.org/10.1136/medethics-2021-107564 PMID: 34610980

32. Bulliard J.-L., Ducros C., Germann S., Arveux P., & Bochud M. (2020). Revue Médicale Suisse: Nouvelles exigences légales dans le domaine de l'enregistrement du cancer: opportunités et défis. *Revue Médicale Suisse*, 16, 2099–2103. https://doi.org/10.53738/REVMED.2020.16.713.2099

33. Johansson A. L., Larønningen S., Skovlund C. W., Kristiansen M. F., Mørch L. S., Friis S., et al. (2022). The impact of the COVID-19 pandemic on cancer diagnosis based on pathology notifications: A comparison across the Nordic countries during 2020. *International Journal of Cancer*, 151, 381–395. https://doi.org/10.1002/ijc.34029 PMID: 35419824

Estimating cancer incidence accounting for changes in cancer detection

Bastien Trächsel, Valentin Rousson, Jean-Luc Bulliard, Isabella Locatelli

1 Introduction

Cancer is an important public health issue in modern societies. In Switzerland, each year, about 45'000 new cases of cancer are diagnosed, and 17'000 deaths are induced [1]. Cancer is the leading cause of death in women aged 25-84 and men aged 45-84 [1]. Analysis of population exposure to carcinogens and behavioral and environmental risk factors is crucial in an epidemiological perspective and for tailoring specific public health interventions. To this end, monitoring annual age-standardized cancer incidence rates may provide a partially distorted view, due to changes in detection over time that may affect observed cancer incidence rates.

Several studies have analyzed the relationship between screening and cancer incidence and have attempted to disentangle underlying incidence trends from the contribution of screening changes, particularly in the context of prostate cancer and the prostate-specific antigen (PSA) screening test. For example, [2] developed a likelihood model for observed incidence, using data on PSA screening frequencies and cancer detection rates over time to simultaneously estimate the mean lead time, i.e. time by which screening advances clinical diagnosis with symptoms [3], and a smooth secular incidence trend. In another study [4], the impact of PSA screening on prostate cancer incidence was investigated

separately for low- and high-stage cancers, based on a simulation model of the natural progression of cancer through pathological stages until clinical (symptomatic) detection [5], also including information on age-specific annual PSA testing frequencies observed in the population. A similar model of prostate cancer natural history [6] was adopted by [7] to predict (past and) future cancer overdiagnosis, i.e. cancers that would not have been detected in the absence of screening, whether the PSA screening program is continued or discontinued at some point. These studies show that screening practices can largely impact the observed trends in cancer incidence.

The association between cancer detection and the cancer stage distribution has also been abundantly examined. For example, using Osaka Cancer Registry data (Japan), [8] examined the stage distribution and the proportion of screen-detected cases for stomach, colorectal, lung, breast and cervical cancer, and found a strong association over time between high proportions of screen-detected cases and high proportions of low stage cancers. A similar result was obtained by [9] based on data from 16 population-based cancer registries (from Belgium, Denmark, England, France, Italy, Ireland, the Netherlands, Slovenia, and Spain), showing significantly higher proportions of low-stage cancers (stages I and II) among screen-detected than non-screen-detected cases.

In the present study, we sought to adjust the observed cancer incidence for changes in cancer detection such as but not limited to: the introduction or modification of screening programs, the use of different screening tools, and incidental detection. Throughout this article, we will loosely refer to changes in cancer detection as changes in 'cancer screening'. Contrary to [4], we did not make any assumptions about the progression of cancer over time. Furthermore, we consider a situation in which data on annual screening frequencies are not readily available, as is the case in Switzerland. Instead, we assumed that in the absence of changes in cancer detection, the distribution of cancer stages among cases remains constant, thus interpreting any change in this distribution

over time as a proxy for a change in detection. This hypothesis, although difficult to test, is however consistent with the results of [8, 9] which show a strong association between screening detection and cancer stage. The paper is organized as follows: the context is described in Section 2; our adjusting model is described in Section 3, along with a classical approach to incidence modeling. Section 4 is devoted to check by simulation the properties of our proposed approach compared to the classical one, and an application to registry data is presented in Section 5. A discussion of our principal findings follows in Section 6.

2 Setting

Cancer is a disease in which certain cells in the body grow uncontrollably and spread to other parts of the body [10]. Most cancers have four stages, determined by a variety of factors, including the size and location of the tumor. Stage I indicates that the cancer is localized in a small area; stage II is cancer that has grown but has not spread. Stage III is cancer that has grown and spread to lymph nodes or other tissues. Stage IV (metastatic or advanced cancer) indicates that the cancer has spread to other organs or areas of the body [11].

In most countries, cancer information is collected through cancer registries. For example, in Switzerland, there are several population-based cancer registries, of which the registry of the canton Geneva is the oldest. It records information on all incident cases of malignant tumors occurring since 1970 in resident population according to international rules, such as the International Classification of Diseases for Oncology [12]. For each new case, information on the date of incidence, the site of the cancer and, for some cancer sites and from 1991 onwards, the stage of the cancer (I-IV) is systematically recorded, among others. It should be noted that a new cancer case only enters the registry on the time of its detection, so that information is only available at that date, for example regarding the stage of the cancer.

The underlying cancer trajectory can be represented by a set of probabilities of transition through the different pathological stages and from each to registration (Figure 1). In particular, $\alpha_{1,t}, .., \alpha_{4,t}$ in Figure 1 are detection (registration) probabilities at each stage in a given year t, driving the distribution of cancer stages that year. For example, the implementation of a thorough screening campaign results in elevated probabilities $\alpha_{1,t}$ and $\alpha_{2,t}$ for detecting cancer at early stages (I or II) through screening, otherwise there is a greater chance of it being clinically detected with symptoms at advanced stages (III or IV). Consider a population of 1 million people. If the cancer risk is constant over time at $\gamma_0 = 1/100$, and all transition probabilities remain constant, the number of cancers recorded each year D_t will be stable around 10'000, whatever the value of detection probabilities $\alpha_{1,t}$, $\alpha_{2,t}$, $\alpha_{3,t}$ and $\alpha_{4,t}$, at least after a certain adjustment period depending on the time of evolution of the tumor. Now consider a sudden change in screening standards at year 31, such as a new screening program, leading to an increase in the probabilities of detection by screening at early stages, for example at times $t = 0, ..., 30$, $\alpha_t = (\alpha_{1,t}, \alpha_{2,t}, \alpha_{3,t}, \alpha_{4,t}) = (0.15, 0.25, 0.25, 0.25)$ and at times $t \geq 31$, $\alpha_t = (0.4, 0.3, 0.25, 0.20)$. The impact of this change on the observed number (incidence) of cancers D_t is illustrated in Figure 2. D_t will temporarily increase, before returning to previous levels within a few years. Thus, a sudden change in screening standards will have only a temporary impact on D_t, as also noted by [4], with the magnitude of the change in detection probabilities at each stage entirely determining the extent of this temporary deviation in observed incidence. Of course, the situation depicted in Figure 2 is quite simplistic, because in practice changes in screening patterns are gradual, even if a new screening standard is suddenly introduced.

The model representing the evolution of the cancer cases in Figure 1 (simplified by removing $\gamma_{0,t}$) is a Markov model with the transition probabilities:

$$P_t = \left[\begin{array}{ccccc|c} x_{1,t} & \gamma_{1,t} & 0 & 0 & \alpha_{1,t} \\ 0 & x_{2,t} & \gamma_{2,t} & 0 & \alpha_{2,t} \\ 0 & 0 & x_{3,t} & \gamma_{3,t} & \alpha_{3,t} \\ 0 & 0 & 0 & x_{4,t} & \alpha_{4,t} \\ \hline 0 & 0 & 0 & 0 & 1 \end{array}\right] = \left[\begin{array}{ccccc} & & & & \alpha_{1,t} \\ & & W_t & & \alpha_{2,t} \\ & & & & \alpha_{3,t} \\ & & & & \alpha_{4,t} \\ \hline 0 & 0 & 0 & 0 & 1 \end{array}\right]. \quad (1)$$

Figure 1 and the associated transition matrix P_t serve as the bases for our simulation study. Here, the incidence D_t will be generated allowing a change in time of the transition matrix P_t. Different situations will be considered, with possibly complex baseline incidence patterns and a progressive introduction of changes in screening habits or improvements in screening techniques leading to a gradual change of matrix P_t. Our proposed model (next section) will be applied to the simulated incidence D_t and results compared with the number of new cancers that would be observed if there were no change in the the transition matrix P_t (no changes in screening) over time. This theoretical quantity, which we call D_t^*, can be obtained explicitly from the transition matrix:

$$D_t^* = \gamma_{0,t-r} \cdot N_{t-r} \quad (2)$$

In equation (2) N_t represent the amount of population at time t , and r is the lag corresponding to the average time required for cancers to progress from stage I to the recorded cancer stage. The last can be obtained as $r = \sum_{j=1}^{4} F_{t(1,j)}$, with $F_t = (I - W_t)^{-1}$.

In the next section, we present the model we will adopt to estimate D_t^*, i.e. to adjust the observed incidence D_t for changes in screening. The model will then be applied to simulated data in Section 6, and to real data from the Geneva cancer registry in Section 7.

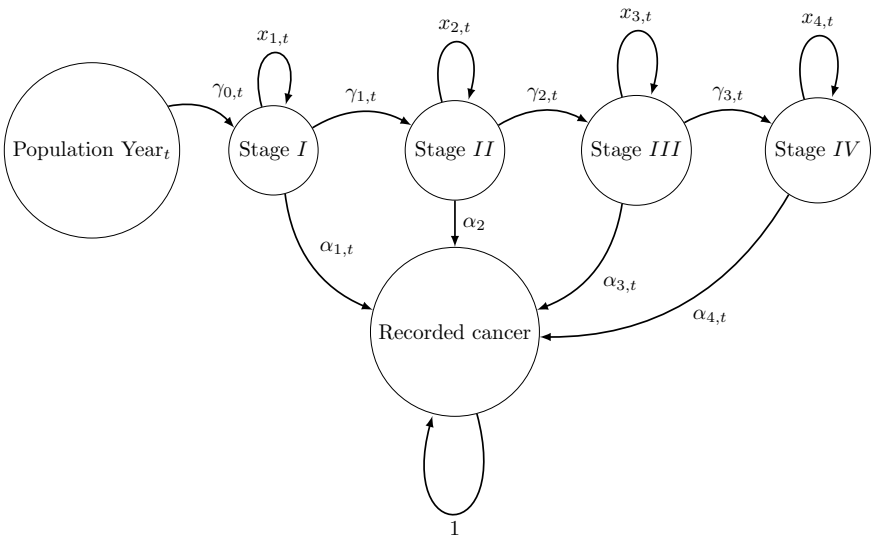

Figure 1: Markov model describing the underlying cancer trajectory through the different pathological stages of cancer and from each to registration

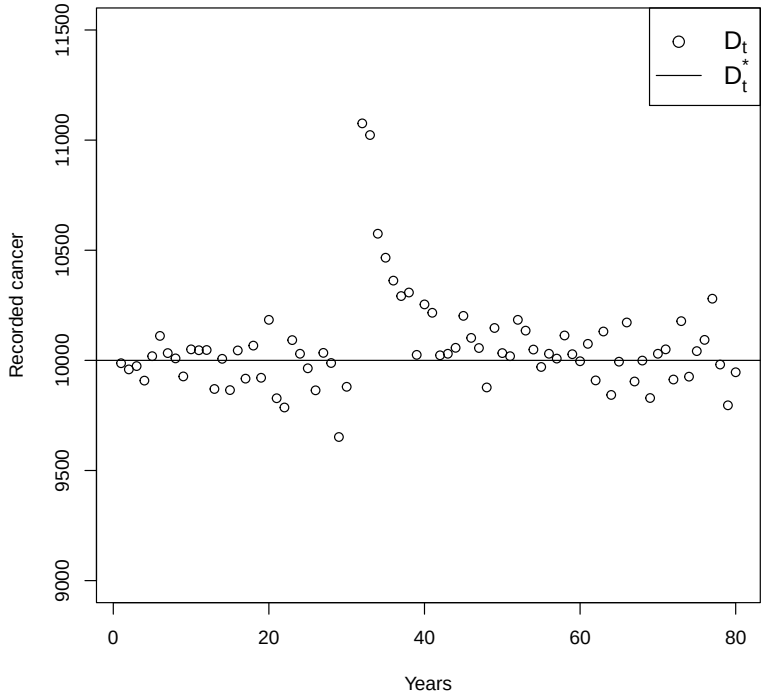

Figure 2: Temporal changes in recorded cancers induced by a sudden change in screening practices

3 Model

3.1 Spline model

Modeling the number of new cancer cases D_t over time is generally based on smoothing functions [13], such as in age-period-cohort (APC) [14, 15, 16] or Bayesian age-period-cohort (BAPC) [17, 18, 19] models, which rely on cubic splines or second-order random walks to estimate the effect of age, time (period), and cohort. As a baseline for comparison, here we adopt a similar, though simpler, approach based on natural cubic splines (ns) for the time only effect:

$$log(D_t) = ns(t) + \varepsilon \tag{3}$$

In the following simulation (see Section 4.1 below) we maintained a constant population to avoid using an offset. We will refer to this model as the NS model.

3.2 Screening Adjusted Model

The model we propose to disentangle underlying trends in cancer incidence from the effect of changes in screening, as approximated by changes in the cancer stage distribution, is an Almond distributed lag model with constraints [20, 21].

Let $f(t)$ define a smooth function of the underlying incidence net of changes in screening (D_t^*), and s_{it} ($i = 1, 2, 3, 4$) the observed proportion of new stage I to IV cancer cases at time t. Information conveyed by s_{it} is a constrained information, since $\sum_{i=1}^{4} s_{it} = 1$. For this reason, these quantities can not be introduced directly into a model. One solution would be using compositional data with isometric logratio transformation [22], projecting data from a four-dimensional to a three-dimensional space. However, since each effect must be lagged several times into the past, this solution would result in too many parameters for a 20-30 year data series.

We opted to approximate the stage distribution s_{it} $(i = 1, 2, 3, 4)$ by its mean $\bar{s}_t$:

$$\bar{s}_t = \frac{1}{4} \sum_{i=1}^{4} i \cdot s_{it} \tag{4}$$

An increase in screening before time t, raising the proportion of stage I cancers in t, s_{1t}, and lowering the proportions of high-stage cancers s_{4t}, will have a negative impact on the mean stage $\bar{s}_t$, and will result in a temporary increase in cancer incidence D_t. Indeed, as seen in previous Section 2, only *changes* in cancer detection (and registration) probabilities α, underlying the observed cancer stage distribution $s_1, ..., s_4$ (and its mean $\bar{s}_t$), and not their absolute value, have an impact on cancer incidence. In addition, a given change has an impact for several years after its occurrence. Therefore, the mean stage $\bar{s}_t$ was differentiated before entering a distributed lag model [20] on the (log-transformed) incidence:

$$log(D_t) = f(t) + \beta_0 \Delta \bar{s}_t + \beta_1 \Delta \bar{s}_{t-1} + ... + \beta_L \Delta \bar{s}_{t-L} + \varepsilon_t. \tag{5}$$

In (5), $\Delta \bar{s}_t$ represents change in mean stage between two consecutive years, $\Delta \bar{s}_t = \bar{s}_t - \bar{s}_{t-1}$, and ε_t are normally distributed errors. Some concerns of this model are the difficulty to select a maximum lag L and the high multicollinearity in the $\bar{s}_t$ [20].

One classical approach to tackle such multicollinearity is imposing some structure on the coefficients of the distributed lag-models [20, 21] β using the Almon polynomial lag model. We assume that the β_l, $l = 1, ..., L$ can be represented as a polynomial in l. For example, we will assume a polynomial of degree 5. This will ensure that the polynomial is flexible enough.

$$\beta_l = \delta_0 + \delta_1 l + \delta_2 l^2 + \delta_3 l^3 + \delta_4 l^4 + \delta_5 l^5 \tag{6}$$

$$log(D_t) = f(t) + \sum_{l=0}^{L}(\delta_0 l^0 + \delta_1 l^1 + \cdots + \delta_5 l^5)\Delta\bar{s}_{i-l} + \varepsilon_t = f(t) + g(t) + \varepsilon_t \quad (7)$$

here the effects are decomposed in $f(t)$ representing the underlying incidence net of changes in screening and $g(t)$ representing the effects of the mean stage changes.

We constrained the β coefficients 1) to be always negative or zero: $\beta_l \leq 0$, $l = 0, .., L$, as a decrease in mean stage should (temporarily) raise incidence, and 2) to be monotonic: $d\beta_l/dl \geq 0$, because the impact on the incidence of a change in mean stage should decrease over time.

The model was estimated via penalized constrained least squares PCLS [23], with starting values set to $\delta_0 = -L - 1$, $\delta_1 = L$ and $\delta_p = 0$, for $p = 2, .., 5$, i.e. the absolute value of the effects decrease linearly over time. We will refer to this model as the Screening Adjusted model (SA).

3.3 Comparison metric

In each simulated scenario (see Section 4.1 below) our screening adjusted model (Section 3.2) will be compared with the classical spline model (Section 3.1) via the absolute distance between the models' predicted counts $\hat{D}_t$ and the theoretical incidence in absence of changes in screening D_t^*:

$$\frac{1}{T}\sum_{t=1}^{T} \mid \log D_t^* - \log \hat{D}_t \mid \quad (8)$$

where T is the length of the time period considered.

4 Simulation

4.1 Setting

In our simulation study we considered a constant population over time. Three different functions for the underlying incidence were considered: flat as in Figure

2 ($\gamma_{0,t}$ constant over time), linear ($\gamma_{0,t} = b \cdot t$) or non-linear ($\gamma_{0,t} = \sin(1000 \cdot \frac{t}{10})$). In addition, we varied the time at which the probabilities begin to transition ($t = \emptyset$, $t = 19$, $t = 26$), and the duration of the transition period between two sets of probabilities α (2 or 6 years). In total, we worked with 15 distinct scenarios (Figure 3). The distance (8) was calculated and averaged over 1000 replications for each scenario.

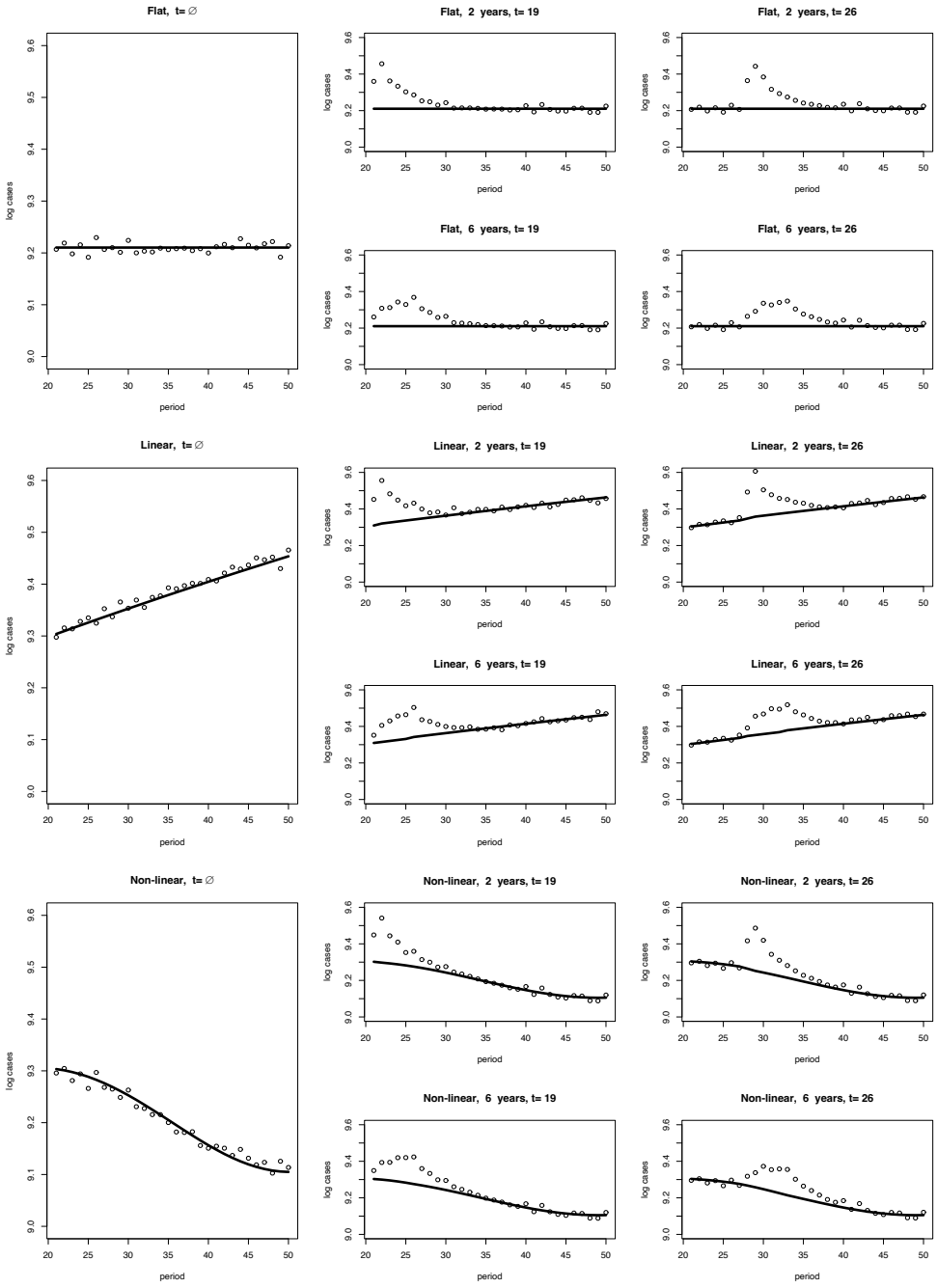

Figure 3: Illustration of one simulation of each of the 15 scenarios. The incidence trends can be either flat, linear or non-linear. The transition time between two sets of registration probabilities can be either 2 years or 6 years, and the time at which the screening starts to change can be either $\emptyset$, i.e. no changes, 19 or 26. Dots are observed numbers of cases, dark lines are numbers of cases which would be observed if there was no change in screening

4.2 Results

First consider as an example one repetition of the last depicted scenario, with a decreasing underlying incidence following a nonlinear curve and a change in the transition matrix P_t that occurs gradually in six periods starting at time $t = 26$. In Figure 4 the dots represent the observed incidence and the black line the theoretical incidence (D_t^*) that would have been observed if there had been no change in the transition probabilities. Predictions of the NS model, not surprisingly, follow the pattern of the observed incidence. In contrast, the SA model manages to correct for the temporary increase in the number of cases and provides a fairly good estimate of the theoretical incidence.

Global results of our simulation study are reported in Table 1. This table contains for each model the mean absolute distance between the predicted and theoretical incidence over the 1000 replicates in each of the 15 scenarios considered, according to 1) the form of the theoretical incidence: flat, linear or non-linear; 2) the time of the change: no change ($t = \emptyset$), change at time 19 (just before the start of the observation period) or during the observation period at time 26; and 3) the length of the transition period between two sets of probabilities: 2 years or 6 years. While the performance of the two models is almost identical when no change in the transition matrix is introduced, it is uniformly better for the SA model in all other scenarios. The fit of the SA model to the theoretical incidence is slightly better when the change occurs during the observation period (t=26) than when it occurs just before the beginning of this period (t=19). On the other hand, the time needed to move from one probability set to another (2 or 6 years) has little impact on the performance of our model.

These results are confirmed in Figure 5 which shows for each model (SA continuous and NS dashed) the distribution of our comparison metric according to the time of the change. If no change occurs (black curves) not only do both models

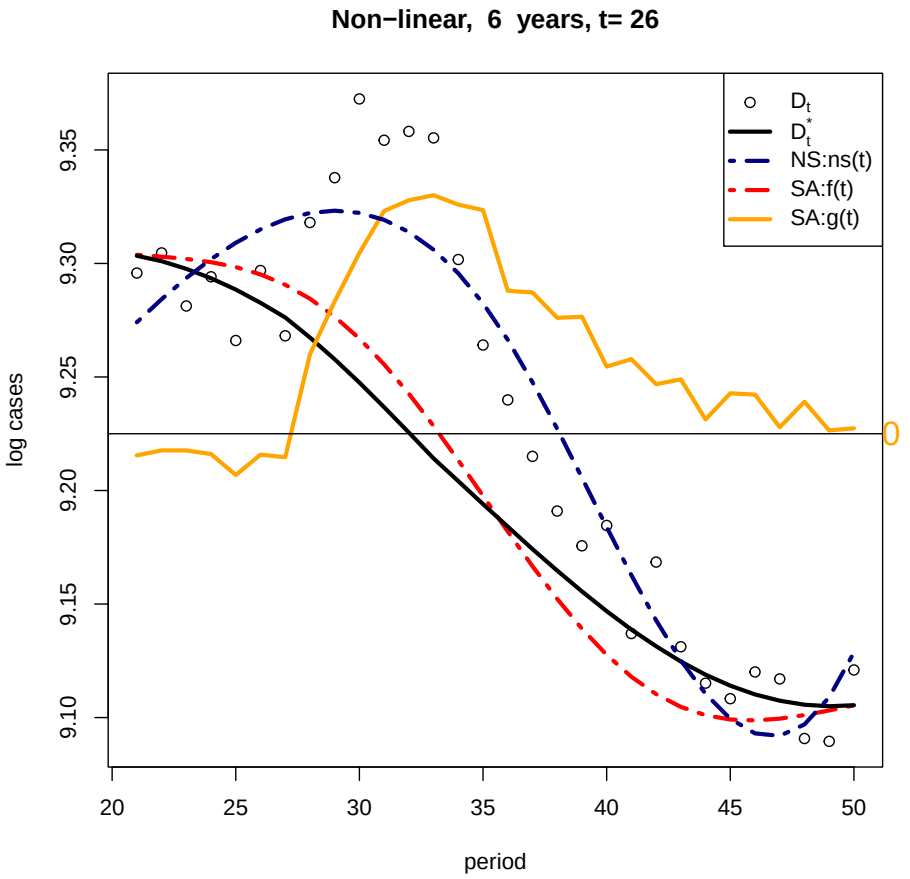

Figure 4: Illustration of the prediction of NS and SA models in one repetition of a selected scenario: non-linear underlying incidence and gradual change in 6 years starting from $t = 26$

estimate the same quantity on average (Table1), but the distribution of results is also almost identical (same variability) for both models. Our model therefore does not perform worse than a classical model when no short-term effects due to changes in screening techniques and practices are observed. On the other hand, the superiority of the SA model over the classical model is evident when it comes to correcting for such effects, even though the correction is not perfect and the results show a higher variability than in the case without screening change (Figure 5, orange and green solid lines).

		t=∅		t=19		t=26	
		SA	NS	SA	NS	SA	NS
Flat	2 years	0.003	0.003	0.017	0.036	0.012	0.042
	6 years			0.014	0.038	0.011	0.042
Increase	2 years	0.005	0.005	0.015	0.035	0.013	0.042
	6 years			0.014	0.038	0.012	0.042
Shift	2 years	0.004	0.005	0.017	0.035	0.010	0.042
	6 years			0.014	0.038	0.012	0.042

Table 1: Mean absolute distance achieved by the screening adjusted (SA) model and the natural spline (NS) model over 1000 replications in each of 15 scenarios

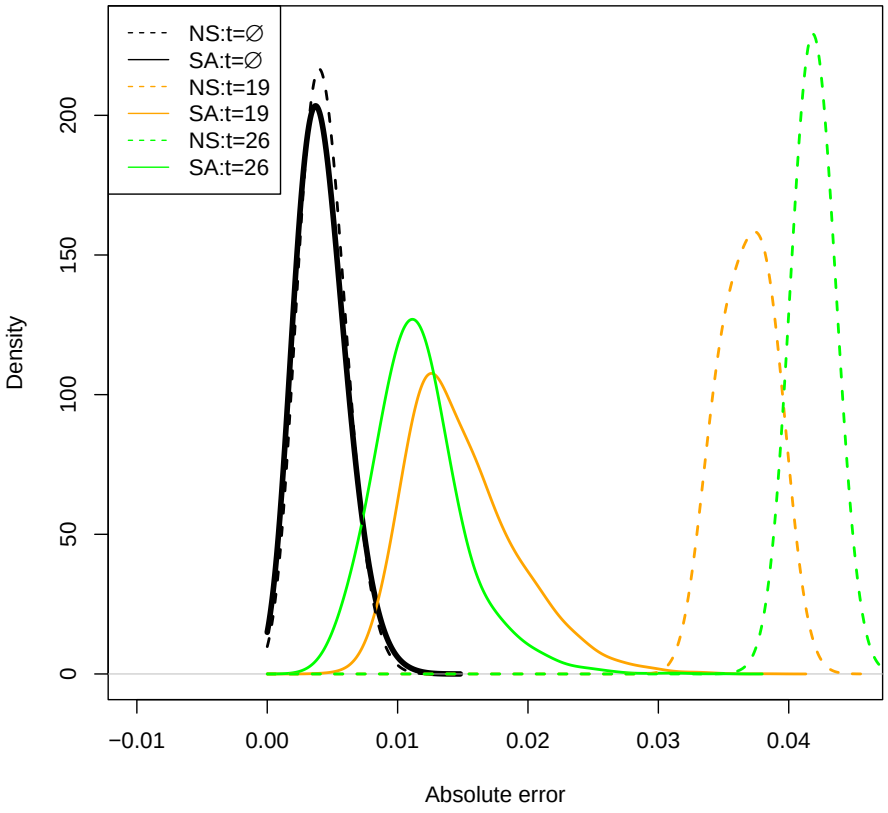

Figure 5: Density of the absolute distances for the NS and SA models according to the time of change (no change $t = \emptyset$, change at $t = 19$ and change at $t = 26$)

5 Application to registry data

The two models presented above were applied to breast cancer incidence data in the canton of Geneva. For these data, a sufficiently long period of stage registration was available, from 1982 to 2016.Because of too high percentages of missing values in this variable for the years between 1982 and 1990 (>10 %), we consider only the period 1991-2016 in our analysis. Entering the stages with 8 lags into the model, this allowed us to obtain estimates from 1999 onwards. The observed incidences and estimates obtained with the NS and SA models are shown in Figure 6 , with an estimate of the effect of the mean stage ($g(t)$) in the SA model (centered at zero). If the latter exceeds zero for a few years, a temporary effect of an increase in screening on the observed incidence is detected and the estimate obtained with the SA model will correct downwards the estimate obtained with the NS model. This is the case for the periods before 2003 and between 2013 and 2015.

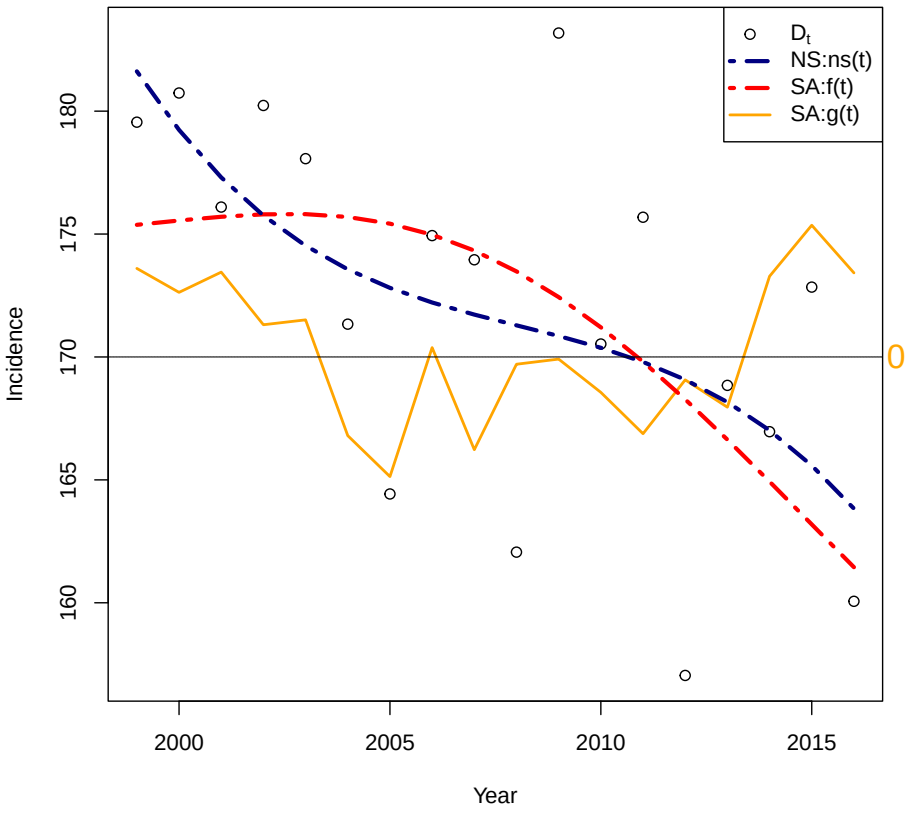

Figure 6: Screening adjusted breast cancer incidence trend in red, unadjusted trend in blue, observed incidence in points, and effects explained by the changes in mean stage (induced by screening changes) in orange (centered). There have been more screening in the years leading 2005, a stabilization in the years leading 2006-2013 and an increase in the years before 2014-2016.

6 Discussion

The present study proposes a new method to correct the observed cancer incidence to take into account changes in cancer detection and thus provide an estimate of the incidence evolution that would have been observed if such changes had not taken place. Unlike previous studies in this field, which make very strong assumptions about the evolution of a certain cancer from one stage to another until detection [2][4] and which are based on knowledge of the evolution over time of the frequency of screening at the population level [6], our model is based on the simple assumption that, in the absence of changes in detection, the distribution of stages detected for a certain cancer will remain approximately constant. Consequently, any change in the distribution of stages, with for example a higher proportion of low-level stages, can be interpreted as an effect of a change in detection, for example an increment in practices due to a new screening policy. This hypothesis is supported by a series of studies that have demonstrated the link between screening and stage distribution [8, 9]. We have thus incorporated the observed changes in the stage distribution, more particularly changes in the mean stage, into a classical log-linear model of cancer incidence as a function of time [20, 21], thus adjusting the incidence estimate for changes in cancer detection (SA model).

In a simulation study considering several scenarios of incidence evolution over time and introduction of screening changes, our model was able to approach the theoretical incidence in the absence of screening changes and thus correct a large part of the bias of the classical estimation, which is highly impacted by these changes, while presenting a performance identical to that of the classical model when no screening changes are introduced in the simulation. On real registry data on breast cancer incidence in the Swiss canton of Geneva, our model corrected for some phases of increased incidence, interpreting them as temporary effects of an increment in screening during the preceding years . The evolution of screening over time in a given context is far from being easily identifiable. In

addition to more or less continuous improvements in screening techniques and policy changes leading to progressive changes in the use of screening, one can also expect random fluctuations due to opportunistic cancer detections or other random variations in the detection process. We believe that our model, without directly considering screening frequencies over time, but simply the evolution of the mean stage among detected cases, is able to smooth out such random fluctuations, thus allowing a more accurate estimation and more robust interpretation of the underlying trend.

The methodology proposed in this paper has also some limitations. The first is to describe the entire cancer stage distribution by its mean. Although the latter is a natural choice when one does not wish to multiply the number of parameters, this simplification leads to some loss of information responsible for the (small) residual bias in our model. A second limitation is that we implicitly assume that increasing screening has no impact on the underlying incidence. This is not true for some cancers when screening detects precancerous lesions and thus prevents the development of a cancer, which is notably the case for colon cancer. Our model does not apply to these situations. A third limitation is related to the availability and quality of registry data, which may limit the applicability of our model. Indeed, the SA model requires that information on stage distribution is available for several periods and that the missing stage remains relatively constant over the periods considered for analysis. In practice, it is often observed that the percentage of missing stages decreases sharply over time, which may result in difficulties in using cancer stages. In addition, the completeness and quality of cancer stage registration may vary over time and across regions and countries. Thus, the application of the SA model may be limited to certain cancer sites or populations for which cancer stage registration is of sufficient quality. However, it should be noted that over time, the completeness and quality of cancer stage registration is very likely to improve, allowing the model to be used for a greater number of cancer sites. Longer and

high-quality incidence series will also allow for a more accurate formulation of the model by introducing the entire stage distribution (through compositional data) instead of its mean.

Despite these limitations, our results suggest that the SA model can be a useful tool for adjusting for the effects of screening in cancer incidence data and can be a valuable addition to the toolkit of cancer researchers and epidemiologists. Its use can help improve our understanding of cancer incidence trends, aid in projections of these trends, and help public health policies be more effective.

References

[1] OFS. Cancer, nouveaux cas et décès: nombre, taux et évolution par localisation cancéreuse et période - 1987-2017 | Tableau, November 2020.

[2] Donatello Telesca, Ruth Etzioni, and Roman Gulati. Estimating Lead Time and Overdiagnosis Associated with PSA Screening from Prostate Cancer Incidence Trends. *Biometrics*, 64(1):10–19, March 2008.

[3] Shan Cheung, Jane L. Hutton, and Julia A. Brettschneider. Review of sojourn time calculation models used in breast cancer screening. Technical Report No. 17-04, Warwick Centre for Research in Statistical Methodology, 2017.

[4] Ruth Etzioni, Roman Gulati, Seth Falcon, and David F. Penson. Impact of PSA Screening on the Incidence of Advanced Stage Prostate Cancer in the United States: A Surveillance Modeling Approach. *Medical Decision Making*, 28(3):323–331, May 2008.

[5] M Cowen. A Markov model of the natural history of prostate cancer. *Journal of Clinical Epidemiology*, 47(1):3–21, January 1994.

[6] A. Tsodikov, A. Szabo, and J. Wegelin. A population model of prostate cancer incidence. *Statistics in Medicine*, 25(16):2846–2866, August 2006. Publisher: John Wiley & Sons, Ltd.

[7] Roman Gulati, Alex Tsodikov, Ruth Etzioni, Rachel A. Hunter-Merrill, John L. Gore, Angela B. Mariotto, and Matthew R. Cooperberg. Expected population impacts of discontinued prostate-specific antigen screening. *Cancer*, 120(22):3519–3526, November 2014. Publisher: John Wiley & Sons, Ltd.

[8] Yasuhiro Toyoda, Takahiro Tabuchi, Hitomi Hama, Toshitaka Morishima, and Isao Miyashiro. Trends in clinical stage distribution and screening

detection of cancer in Osaka, Japan: Stomach, colorectum, lung, breast and cervix. *PLOS ONE*, 15(12):e0244644, December 2020.

[9] Rafael Cardoso, Feng Guo, Thomas Heisser, Harlinde De Schutter, Nancy Van Damme, Mef Christina Nilbert, Jane Christensen, Anne-Marie Bouvier, Véronique Bouvier, Guy Launoy, Anne-Sophie Woronoff, Mélanie Cariou, Michel Robaszkiewicz, Patricia Delafosse, Florence Poncet, Paul M. Walsh, Carlo Senore, Stefano Rosso, Valery E.P.P. Lemmens, Marloes A.G. Elferink, Sonja Tomšič, Tina Žagar, Arantza Lopez de Munain Marques, Rafael Marcos-Gragera, Montse Puigdemont, Jaume Galceran, Marià Carulla, Antonia Sánchez-Gil, María-Dolores Chirlaque, Michael Hoffmeister, and Hermann Brenner. Overall and stage-specific survival of patients with screen-detected colorectal cancer in European countries: A population-based study in 9 countries. *The Lancet Regional Health - Europe*, 21:100458, October 2022.

[10] NCI U.S. National Cancer Institute. What Is Cancer? - NCI, September 2007. Archive Location: nciglobal,ncienterprise.

[11] NIH U.S. National Cancer Institute. Tumor Grade - NCI, May 2013. Archive Location: nciglobal,ncienterprise.

[12] April Fritz, Constance Percy, Andrew Jack, Kanagaratnam Shanmugaratnam, Leslie H. Sobin, D. Maxwell Parkin, Sharon L. Whelan, and World Health Organization. International classification of diseases for oncology. *Classification internationale des maladies pour l'oncologie*, 2000. Edition: 3rd ed Place: Geneva Publisher: World Health Organization Section: English edtion reprinted in 2001,2005, 2007, 2009, 2012.

[13] Xue Qin Yu, Qingwei Luo, Suzanne Hughes, Stephen Wade, Michael Caruana, Karen Canfell, and Dianne L O'Connell. Statistical projection methods for lung cancer incidence and mortality: a systematic review. *BMJ Open*, 9(8):e028497, August 2019.

[14] Bendix Carstensen, Martyn Plummer, Esa Laara, and Michael Hills. *Epi: A Package for Statistical Analysis in Epidemiology.* CRAN, 2022.

[15] Theodore R. Holford. The Estimation of Age, Period and Cohort Effects for Vital Rates. *Biometrics*, 39(2):311–324, 1983. Publisher: [Wiley, International Biometric Society].

[16] Mark J. Rutherford, John R. Thompson, and Paul C. Lambert. Projecting Cancer Incidence using Age-period-cohort Models Incorporating Restricted Cubic Splines. *The International Journal of Biostatistics*, 8(1), January 2012.

[17] Andrea Riebler, Leonhard Held, and Håvard Rue. Estimation and extrapolation of time trends in registry data—Borrowing strength from related populations. *The Annals of Applied Statistics*, 6(1), March 2012.

[18] Andrea Riebler and Leonhard Held. Projecting the future burden of cancer: Bayesian age-period-cohort analysis with integrated nested Laplace approximations: Projecting the future burden of cancer. *Biometrical Journal*, 59(3):531–549, May 2017.

[19] Volker Schmid and Leonhard Held. Bayesian Extrapolation of Space-Time Trends in Cancer Registry Data. *Biometrics*, 60(4):1034–1042, 2004. Publisher: [Wiley, International Biometric Society].

[20] G. S. Maddala. *Econometrics*. Economics handbook series. McGraw-Hill, New York, 1977.

[21] Shirley Almon. The Distributed Lag Between Capital Appropriations and Expenditures. *Econometrica*, 33(1):178, January 1965.

[22] John Aitchison. Principles of Compositional Data Analysis. *Lecture Notes-Monograph Series*, 24:73–81, 1994. Publisher: Institute of Mathematical Statistics.

[23] Simon N. Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 2 edition, May 2017.