



# Quantifying the Predictive Accuracy of a Polygenic Risk Score for Predicting Incident Cancer Cases : Application to the CARTaGENE Cohort

Julianne Duhazé<sup>1,2\*</sup>, Rodolphe Jantzen<sup>1,3</sup>, Yves Payette<sup>3</sup>, Thibault De Malliard<sup>3</sup>, Catherine Labbé<sup>3</sup>, Nolwenn Noisel<sup>3</sup> and Philippe Broët<sup>1,2,3\*</sup>

## OPEN ACCESS

### Edited by:

Celia M. T. Greenwood,  
Lady Davis Institute (LDI), Canada

### Reviewed by:

Paramita Saha-Chaudhuri,  
McGill University, Canada  
Shelley B. Bull,  
Lunenfeld-Tananbaum Research  
Institute of Mount  
Sinai Hospital, Canada

### \*Correspondence:

Julianne Duhazé  
julianne.joseph-duhaze@u-psud.fr  
Philippe Broët  
philippe.broet@inserm.fr

### Specialty section:

This article was submitted to  
Statistical Genetics and Methodology,  
a section of the journal  
Frontiers in Genetics

**Received:** 02 December 2020

**Accepted:** 31 March 2020

**Published:** 24 April 2020

### Citation:

Duhazé J, Jantzen R, Payette Y,  
De Malliard T, Labbé C, Noisel N and  
Broët P (2020) Quantifying the  
Predictive Accuracy of a Polygenic  
Risk Score for Predicting Incident  
Cancer Cases : Application to the  
CARTaGENE Cohort.  
*Front. Genet.* 11:408.  
doi: 10.3389/fgene.2020.00408

<sup>1</sup> Research Center, CHU Sainte-Justine, Montreal, QC, Canada, <sup>2</sup> University Paris-Saclay, CESP, INSERM, Villejuif, France, <sup>3</sup> CARTaGENE, CHU Sainte-Justine, Montreal, QC, Canada

With the increasing use of polygenic risk scores (PRS) there is a need for adapted methods to evaluate the predictivity of these tools. In this work, we propose a new pseudo- $R^2$  criterion to evaluate PRS predictive accuracy for time-to-event data. This new criterion is related to the score statistic derived under a two-component mixture model. It evaluates the effect of the PRS on both the propensity to experience the event and on the dynamic of the event among the susceptible subjects. Simulation results show that our index has good properties. We compared our index to other implemented pseudo- $R^2$  for survival data. Along with our index, two other indices have comparable good behavior when the PRS has a non-null propensity effect, and our index is the only one to detect when the PRS has only a dynamic effect. We evaluated the 5-year predictivity of an 18-single-nucleotide-polymorphism PRS for incident breast cancer cases on the CARTaGENE cohort using several pseudo- $R^2$  indices. We report that our index, which summarizes both a propensity and a dynamic effect, had the highest predictive accuracy. In conclusion, our proposed pseudo- $R^2$  is easy to implement and well suited to evaluate PRS for predicting incident events in cohort studies.

**Keywords:** pseudo- $R^2$ , polygenic risk score, survival models, survival mixture model, breast cancer

## 1. INTRODUCTION

With the power of genotyping technologies, genome-wide association studies (GWAS) focusing on complex diseases have identified a large number of genetic variants associated with various traits of interest (e.g., diabetes, cardiovascular diseases, cancer...) (Buniello et al., 2019). These large GWAS have provided various lists of disease-related single nucleotide polymorphisms (SNP) together with their effect size estimates (McCarthy et al., 2008). Based on these findings, there has been a growing interest recently in deriving polygenic risk scores (PRS) in order to provide individual risk

predictions for various phenotypes (Torkamani et al., 2018). Broadly speaking, for each individual, a classical polygenic score consists of a linear combination of the trait-associated alleles carried by the subject and weighted by their effect sizes. The list of these risk alleles and their corresponding weights are obtained from published GWAS. Keeping in mind classical modeling assumptions of linearity, additivity, and lack of interactions, these PRS provide some estimates of the probabilistic susceptibility of an individual to experience the disease and can be used for disease risk stratification (Torkamani et al., 2018). As a result, in recent years, a burgeoning literature has started to focus on the evaluation of published PRS (for a few: International Multiple Sclerosis Genetics Consortium et al., 2010; Machiela et al., 2011; Khera et al., 2018).

The predictive accuracy of these PRS are usually assessed by measures such as the coefficient of determination (noted as  $R^2$ ) which is interpreted as the proportion of variation in the phenotype that is explained by the PRS. For continuous outcomes, the coefficient of determination is well-defined and unique, however its extension to other outcomes such as binary outcomes is not straightforward. By analogy with the linear model and from different perspectives, various generalizations of the  $R^2$  have been proposed (e.g., Hu et al., 2006). They are usually referred to as pseudo- $R^2$ .

In practice, most of the pseudo- $R^2$  that are used for assessing the predictive accuracy of PRS focus on binary outcomes. This is the case for cancer susceptibility where most of the PRS studies have analyzed cancers as a binary outcome (affected/not affected), irrespective of the age of onset. However, as the occurrence of cancers is strongly influenced by age, prediction modeling should not neglect the dynamic of cancer occurrence over time. Thus, the assessment of predictive accuracy of PRS for cancer susceptibility based upon epidemiological cohorts should use time-to-event data and provide predictivity for a defined risk projection interval.

For such quantification with time-to-event data, a large spectrum of pseudo- $R^2$  have been proposed, most of them relying on the classical Cox proportional hazards model (Cox, 1972). These pseudo- $R^2$  fall mainly into the categories of explained randomness (entropy-based) or explained variation (variance-based). This latter framework corresponds to the proportion of the outcome variance that is explained by the studied covariates. The estimates rely either on comparing empirical survival functions with and without covariates or on statistical quantities which are directly or indirectly related to the likelihood function (O'Quigley, 2008; Flandre et al., 2017). In practice, most of these pseudo- $R^2$  have to maximize the log (partial) likelihood of the full model. However, it is sometimes not a straightforward issue, particularly for complex non-proportional survival models.

We had to face such issue in a recent study relying upon the Quebec population-based cohort CARTaGENE where our main objective was to evaluate the 5-year predictivity of a published PRS for breast cancer. To assess the 5-year predictivity, we had to rely upon a non-standard survival model that

considers age as the time scale and takes into account that a proportion of the individuals are not susceptible to develop the disease within 5 years. Such survival model belongs to the class of survival mixture models where the population under study is a mixture of individuals with those at risk for experiencing the event within 5 years and those who are not at risk (Maller and Zhou, 1996). The motivation behind the use of this survival model is that at a specified age and over a 5-year horizon a woman can be either susceptible or non-susceptible to experience breast cancer and her PRS may be related to the propensity for experiencing the event. Moreover, among the susceptible ones, some of them may experience the event earlier than others and their PRS may be linked to the dynamic of the event. In this context, we know that the classical survival models are not well suited for quantifying both effects.

This issue prompted us to derive a pseudo- $R^2$  which relies on time-to-event data and quantifies the 5-year predictivity accuracy of PRS. This new criterion extends a previous work (Rouam et al., 2010) on pseudo- $R^2$  that was restricted to classical proper survival models. The score statistic derived from the partial likelihood under an entry-age-stratified age-scaled working survival model enables the calculation of the pseudo- $R^2$ . This criterion quantifies the predictive ability of the studied factor to separate subject outcomes on both the probability of experiencing the event and the occurrence dynamic.

In this work, we present this new criterion and report the results obtained from a simulation study. We show its practical interest for evaluating the accuracy of published PRS for predicting incident breast cancer cases in the following 5 years using the Quebec population-based cohort CARTaGENE.

## 2. MATERIALS AND METHODS

### 2.1. Materials: The CARTaGENE Study

CARTaGENE ([www.cartagene.qc.ca](http://www.cartagene.qc.ca)) is a population cohort consisting of 43,037 Quebec residents aged between 40 and 69 years at recruitment (Awadalla et al., 2013). Enrollment of participants began in July 2009 and was carried out in two phases in six metropolitan areas. On their enrollment date, each participant filled out a questionnaire about health, lifestyle, individual, and familial history of disease, and prescribed medication. Only the women were included in the breast cancer study ( $n = 23,797$ ).

Linkage with administrative databases was included in the participant consent form: 1/the Quebec administrative health database MED-ÉCHO contains hospitalizations, claims, and date of death of insured patients (about 98% of Quebec residents) (data available from January 1st, 1998 to March 31st, 2016); 2/the Quebec Breast Cancer Registry contains information about the Quebec Breast Cancer Screening Program, such as mammogram results and breast cancers histological confirmation (data available from May 15th, 1998 to December 31st, 2017).

To comprehensively define women with a breast cancer (invasive or *in situ*), we used an algorithm based on a

previous report from the Institut National de Santé Publique du Québec (INSPQ) (Théberge et al., 2003). Using the Breast Cancer Registry, we retrieved the incidence date of histologically confirmed breast cancers. Then, we selected all women having an abnormal mammography and retrieved, if available, the incidence date from the MED-ÉCHO database for women with at least two claims in 2 years or one hospitalization with the corresponding ICD-9 (174, 2330) and ICD-10 (C50, D05) codes.

For the breast cancer study, only the women with genetic data available and who did not have a breast cancer before their enrollment in the CARTaGENE cohort were selected ( $n = 4,554$ ).

## 2.2. Polygenic Risk Scores

Based on the study of Evans et al. (2017), we computed a PRS for each woman. This PRS uses 18 SNP that have been shown to be associated with breast cancer risk in general European populations together with the published per-allele ORs. Each PRS is the linear combination of the number of risk alleles for each of the 18 SNP weighted by their corresponding log odds-ratios.

Genotyping data has been generated through different projects and using different chips and platforms (Illumina Omni 2.5M, Illumina Infinium Global Screening Array, and Affymetrix Axiom UK biobank). Among the 18 SNP composing the Evans PRS, three were not available in our study and nine had missing data. To impute the missing SNP, we used the Michigan Imputation Server with the Minimac4 algorithm (Das et al., 2016). Imputation reference panel was the HRC r1.1 2016 European population, and the phasing was performed with Eagle v2.4 (Loh et al., 2016). A genetic quality control (QC) was made before the imputation. After imputation, QC was performed on individuals based on the Anderson et al. protocol (Anderson et al., 2010), using all individuals genotyped with the Illumina Infinium Global Screening Array. Individuals with a call rate lower than 95% and a heterozygosity higher than three standard deviations were removed. In pairs of individuals with an identity by state (IBD) higher than 0.1875, the individual with the lowest call rate was removed. To remove participants with divergent European ancestries, we used the first two principal components with the HapMap phase 3 reference panel (The International HapMap 3 Consortium, 2010).

All the QC procedures were performed using PLINK (Purcell et al., 2007) and R package Gaston (Perdry and Dandine-Roulland, 2019).

## 2.3. Outcomes

The outcome was the age at occurrence of breast cancer. Patients without breast cancer occurrence were censored. Censoring time was age at the end of the 5-year study period (administrative censoring) or age at death.

For taking into account the age effect, our survival model was stratified by age at cohort entry with four groups: [39–50], [50–55], [55–60], [60–69].

## 2.4. Methods: The Pseudo-R<sup>2</sup> Criterion

### 2.4.1. Notations and Survival Model

In this work, we considered a survival model with age as the time scale and stratification on age at entry in the study. Let  $A_0$

denote a random variable that corresponds to the age at which the individual free of disease enters the cohort. Let  $A$  be the age at which the individual is experiencing the event of interest with  $A = A_0 + T^*$  where  $T^*$  is the event time (i.e., the time elapsed between the enrollment in the cohort and the date of event). Let  $C$  be the age at which the individual is censored with  $C = A_0 + C^*$  where  $C^*$  is the censoring time (i.e., the time elapsed between the enrollment in the cohort and the date of analysis or last follow-up). Conditionally on  $A_0$ , let the random variables  $A$  and  $C$  assumed to satisfy the condition of independent censoring (Fleming and Harrington, 2005).

In the following, we consider  $J$  strata for age at entry. For a subject  $i$  in stratum  $j$  ( $j = 1, \dots, J$ ), let  $X_{ij} = \min(A_{ij}, C_{ij})$  be the observed follow-up time on the age scale,  $\delta_{ij} = \mathbf{1}_{(X_{ij}=A_{ij})}$  the indicator of event and  $Y_{ij}(s) = \mathbf{1}_{(X_{ij} \geq s)}$  the indicator of being at risk for the event at age  $s$ .  $Y_{ij}(s) = 1$  indicates that subject  $i$  in stratum  $j$  is at risk just before time  $s$ ,  $Y_{ij}(s) = 0$  otherwise. Let  $Z_{ij}$  be the value of the PRS computed for individual  $i$  in stratum  $j$ . For each individual, the observed data consists of  $(X_{ij}, \delta_{ij}, A_{0ij}, Z_{ij})$ .

Since we focus on a 5-year projection interval, we have to take into account that some individuals who enter the cohort are non-susceptible subjects for the event of interest whereas the other are susceptible subjects who may experience the event by the end of the 5 years or may be censored prior to experiencing the event.

Thus, we consider the following two-part semi-parametric multiplicative model whose marginal survival distribution is expressed as:

$$S(a|j, Z = z) = \exp[-\theta_j e^{\alpha z}] + \left\{ 1 - \exp[-\theta_j e^{\alpha z}] \right\} \times \exp[-\Lambda_{0j}(a)e^{\beta z}] \quad (1)$$

where  $\exp[-\theta_j e^{\alpha z}]$  is the probability of being a non-susceptible individual for a subject in stratum  $j$ . This latter quantity (sometimes called the tail defect) depends upon the age at entry and the PRS. Here,  $\theta_j > 0$  is an unknown positive age-stratum parameter and  $\alpha$  an unknown parameter of interest which quantifies the impact of the PRS on the propensity of being a non-susceptible individual.

In this model,  $\exp[-\Lambda_{0j}(a)e^{\beta z}]$  is the conditional survival distribution of time-to-event for stratum  $j$  for those who are susceptible to experience the event within the projection interval. Here,  $\Lambda_{0j}(a)$  is an unspecified conditional baseline cumulative hazard function for stratum  $j$  and  $\beta$  is an unknown parameter of interest which quantifies the impact of the PRS on the dynamic of the event's occurrence.

In the following, we re-write the survival model presented above in terms of the multiplicative hazard functions as:

$$h(a|j, Z = z) = \lambda_{0j}(a)(1 - e^{-\theta_j e^{\alpha z}})e^{\beta z} e^{-\Lambda_{0j}(a)e^{\beta z}} e^{-\log(S(a|j,z))}$$

where  $\lambda_{0j}(a)$  is the first derivative of  $\Lambda_{0j}(a)$ .

From this multiplicative model, a partial likelihood can be written as follows:

$$L(\alpha, \beta, \theta_j, \Lambda_{0j}) = \prod_{j=1}^J \left\{ \prod_{i=1}^{n_j} \left[ \frac{(1 - e^{-\theta_j e^{\alpha z_i}}) e^{\beta z_i} e^{-\Lambda_{0j}(a_i) e^{\beta z_i}} e^{-\log(S_j(a_i|z_i))}}{\sum_{m=1}^{n_j} Y_m(a_i) (1 - e^{-\theta_j e^{\alpha z_m}}) e^{\beta z_m} e^{-\Lambda_{0j}(a_i) e^{\beta z_m}} e^{-\log(S_j(a_i|z_m))}} \right]^{\delta_i} \right\}$$

with  $n_j$  the number of individuals in stratum  $j$  and  $N = n_1 + \dots + n_j$  the total number of individuals. Here,  $S_j(a)$  and  $\Lambda_{0j}(a)$  are the marginal survival function and the conditional baseline cumulative hazard function for stratum  $j$ , respectively.

### 2.4.2. Scores Components

From what precedes, we can easily obtain the two components of the log partial likelihood score function evaluated under the null hypothesis of no PRS effect  $\mathcal{H}_0 : \beta = \alpha = 0$ :

$$U_{1ij} = \frac{\partial LL_{ij}}{\partial \alpha} = \delta_i w_1(a_i) \left\{ z_i - \frac{\sum_{l \in \mathcal{R}_{ij}} Y_l(a_i) w_1(a_i) z_l}{\sum_{l \in \mathcal{R}_{ij}} Y_l(a_i) w_1(a_i)} \right\}$$

$$U_{2ij} = \frac{\partial LL_{ij}}{\partial \beta} = \delta_i w_2(a_i) \left\{ z_i - \frac{\sum_{l \in \mathcal{R}_{ij}} Y_l(a_i) w_2(a_i) z_l}{\sum_{l \in \mathcal{R}_{ij}} Y_l(a_i) w_2(a_i)} \right\}$$

where  $\mathcal{R}_{ij}$  is the risk set of stratum  $j$  including the individual that experienced the event at time  $a_i$ . Here  $w_1(s) = e^{-\theta_j} \theta_j S_j^{-1}(s)$  and  $w_2(s) = \left[ e^{-\theta_j} (1 - e^{-\Lambda_{0j}(s)} - \Lambda_{0j}(s)) + e^{-\Lambda_{0j}(s)} \right] S_j^{-1}(s)$  are the weights for the two components, respectively. The nuisance parameters,  $e^{-\theta_j}$ ,  $\Lambda_{0j}$  and  $S_j(s)$  are the tail defect, the conditional cumulative hazard function and the marginal survival distribution for stratum  $j$  computed under the null hypothesis.

From a biological perspective, the quantity  $U_{1ij}$  can be linked to differences in the propensity of being a non-susceptible individual and  $U_{2ij}$  to differences in the dynamic of the occurrence of the event of interest for susceptible individuals.

### 2.4.3. Scores as Measures of Separability

Following a previous work (Rouam et al., 2010), the non-null quantities  $U_{1ij}$  and  $U_{2ij}$  computed at event time can be reformulated as two measures of separability that quantify the ability of the PRS to separate individuals from the stratum  $j$  who experience the event at time  $a_i$  from those who are still at risk.

The first quantity  $U_{1ij}$  can be re-written as:

$$U_{1ij} = \delta_i w_1(a_i) \left( \frac{\sum_{l \in \mathcal{R}_{ij}^*} Y_l(a_i) w_1(a_i)}{\sum_{l \in \mathcal{R}_{ij}} Y_l(a_i) w_1(a_i)} \right) \times \left\{ z_i - \frac{\sum_{l \in \mathcal{R}_{ij}^*} Y_l(a_i) w_1(a_i) z_l}{\sum_{l \in \mathcal{R}_{ij}^*} Y_l(a_i) w_1(a_i)} \right\}$$

$$U_{1ij} = \delta_i w_1(a_i) \left( \frac{\sum_{l \in \mathcal{R}_{ij}^*} Y_l(a_i) w_1(a_i)}{\sum_{l \in \mathcal{R}_{ij}} Y_l(a_i) w_1(a_i)} \right) \times (z_i - \bar{z}_{w_1})$$

In like manner, the second one is as:

$$U_{2ij} = \delta_i w_2(a_i) \left( \frac{\sum_{l \in \mathcal{R}_{ij}^*} Y_l(a_i) w_2(a_i)}{\sum_{l \in \mathcal{R}_{ij}} Y_l(a_i) w_2(a_i)} \right) \times (z_i - \bar{z}_{w_2})$$

where  $\mathcal{R}_{ij}^*$  is the risk set of stratum  $j$  without the individual that experienced the event at time  $a_i$ .

The two non-null components  $U_{1ij}$  and  $U_{2ij}$  can be interpreted as a weighted differences between the mean value of the genomic score for the individuals who experience the event at time  $a_i$  and the weighted mean for those who are still at risk (i.e., the mixture of individuals who do not experience the event at time  $a_i$ ). Differences close to zero indicate a weak or null separability. Large differences indicate that the individuals are well-separated.

### 2.4.4. Shifted Score Components

In the following, we introduce the shifted scores (or robust scores)  $W_{1ij}$  and  $W_{2ij}$  derived from the seminal work of Lin and Wei (1989). The shifts take into account the dependence between the individuals scores  $U_{1ij}$  and  $U_{2ij}$ . The shifted scores  $W_{1ij}$  and  $W_{2ij}$  are independent and identically distributed (Lachin, 2011) and are as:

$$W_{1ij} = U_{1ij} - \tilde{U}_{1ij} = \delta_i w_1(a_i) \left( z_i - \frac{\sum_{l \in \mathcal{R}_{ij}} Y_l(a_i) w_1(a_i) z_l}{\sum_{l \in \mathcal{R}_{ij}} Y_l(a_i)} \right) - \sum_{l=1}^{n_j} \frac{Y_l(t_l) \delta_l w_1(a_l)}{\sum_{r \in \mathcal{R}_{lj}} Y_r(a_l)} \left( z_i - \frac{\sum_{r \in \mathcal{R}_{lj}} Y_r(a_l) w_1(a_l) z_r}{\sum_{r \in \mathcal{R}_{lj}} Y_r(a_l)} \right)$$

$$W_{2ij} = U_{2ij} - \tilde{U}_{2ij} = \delta_i w_2(a_i) \left( z_i - \frac{\sum_{l \in \mathcal{R}_{ij}} Y_l(a_i) w_2(a_i) z_l}{\sum_{l \in \mathcal{R}_{ij}} Y_l(a_i)} \right) - \sum_{l=1}^{n_j} \frac{Y_l(t_l) \delta_l w_2(a_l)}{\sum_{r \in \mathcal{R}_{lj}} Y_r(a_l)} \left( z_i - \frac{\sum_{r \in \mathcal{R}_{lj}} Y_l(a_i) w_2(a_l) z_r}{\sum_{r \in \mathcal{R}_{lj}} Y_r(a_l)} \right)$$

The practical expressions of these latter quantities are obtained by plugging the two estimates  $\hat{w}_1(s)$  and  $\hat{w}_2(s)$  where we replace  $S_j(s)$  by the left-continuous version of the Fleming-Harrington estimator obtained under  $\mathcal{H}_0$  using the Nelson-Aalen estimator. The nuisance parameter  $\theta$  is estimated by  $\hat{\theta}_j = -\log(1 - S_j(t_{\max}))$  where  $t_{\max}$  is the last observed failure time. The shifts  $\tilde{U}_{1ij} = \hat{\mathbb{E}}(\hat{U}_{1ij})$  and  $\tilde{U}_{2ij} = \hat{\mathbb{E}}(\hat{U}_{2ij})$  are weighted average of the score calculated at times  $s$  prior to time  $a_i$ .

In the following, as we focus on separability measures we will consider only the shifted score components associated with event times that we denoted as  $W_{1ij}^*$  and  $W_{2ij}^*$ .

### 2.4.5. Pseudo- $R^2$ Criterion

In the following we derive a pseudo- $R^2$  criteria which is interpreted as the proportion of variation of separability that is explained by the PRS. Since the classical score contributions ( $U_{ij}$ ) are based on a partial rather than a full likelihood, they are not independently and identically distributed. However, the shifted score contributions ( $W_{ij}$ ) introduced by Lin and

Wei (1989) are independent and identically distributed. As our pseudo- $R^2$  quantifies the proportion of variance explained by the PRS, it relies on variance estimates and this latter condition is important for estimation purposes. Moreover, since our individual score contributions can be expressed as differences between the means of the PRS of the group of patients observed experiencing the event of interest and the group of those observed not experiencing the event, the considered shifted score contributions are those calculated at each occurrence of the event.

We recall that the quantities  $W_{1ij}^*$  and  $W_{2ij}^*$  represent the measures of separability that are calculated at an event time. We denote  $k_j$  the number of event times for stratum  $j$ . The total number of events is denoted by  $K$ . In practice, we use the generalized variance of the two-dimensional random vector  $W = (W_1^*, W_2^*)$  that is defined as the determinant of its variance-covariance matrix. Then, we derive a pseudo- $R^2$  which is based on the relative difference between an estimate of the variance-covariance matrix of the shifted scores computed under the null hypothesis ( $\mathcal{H}_0$ : no effect of the PRS;  $\mathbb{E}[W] = 0$ ) and an estimate calculated under the alternative hypothesis ( $\mathcal{H}_1$ : effect of the PRS;  $\mathbb{E}[W] \neq 0$ ).

Under the null hypothesis ( $\mathcal{H}_0 : \alpha = \beta = 0$ ), we have:

$$\begin{aligned} \hat{\sigma}^2(W_1^*) &= \frac{1}{K-1} \sum_{j=1}^J \left( \sum_{i=1}^{k_j} W_{1ij}^{*2} \right) \\ \hat{\sigma}(W_{12}^*) &= \frac{1}{K-1} \sum_{j=1}^J \left( \sum_{i=1}^{k_j} W_{1ij}^* W_{2ij}^* \right) \\ \hat{\sigma}^2(W_2^*) &= \frac{1}{K-1} \sum_{j=1}^J \left( \sum_{i=1}^{k_j} W_{2ij}^{*2} \right). \end{aligned}$$

Then (omitting the term  $\frac{1}{K-1}$ ) we have :

$$\begin{aligned} \det(\Sigma) &= \sum_{j=1}^J \left( \sum_{i=1}^{k_j} W_{1ij}^{*2} \right) \sum_{j=1}^J \left( \sum_{i=1}^{k_j} W_{2ij}^{*2} \right) \\ &\quad - \left( \sum_{j=1}^J \left( \sum_{i=1}^{k_j} W_{1ij}^* W_{2ij}^* \right) \right)^2 \end{aligned}$$

Under the alternative hypothesis ( $\mathcal{H}_1$ ), we have:

$$\begin{aligned} \hat{\sigma}^2(W_1^*) &= \frac{1}{K-1} \sum_{j=1}^J \left( \sum_{i=1}^{k_j} W_{1ij}^{*2} - \frac{1}{k_j} \sum_{i=1}^{k_j} (W_{1ij}^*)^2 \right) \\ \hat{\sigma}(W_{12}^*) &= \frac{1}{K-1} \sum_{j=1}^J \left( \sum_{i=1}^{k_j} W_{1ij}^* W_{2ij}^* - \frac{1}{k_j} \left\{ \sum_{i=1}^{k_j} W_{1ij}^* \sum_{i=1}^{k_j} W_{2ij}^* \right\} \right) \\ \hat{\sigma}^2(W_2^*) &= \frac{1}{K-1} \sum_{j=1}^J \left( \sum_{i=1}^{k_j} W_{2ij}^{*2} - \frac{1}{k_j} \sum_{i=1}^{k_j} (W_{2ij}^*)^2 \right). \end{aligned}$$

Then:

$$\begin{aligned} \det(\Sigma^*) &= \sum_{j=1}^J \left( \sum_{i=1}^{k_j} W_{1ij}^{*2} - \frac{1}{k_j} \left( \sum_{i=1}^{k_j} W_{1ij}^* \right)^2 \right) \\ &\quad \times \sum_{j=1}^J \left( \sum_{i=1}^{k_j} W_{2ij}^{*2} - \frac{1}{k_j} \left( \sum_{i=1}^{k_j} W_{2ij}^* \right)^2 \right) \\ &\quad - \left( \sum_{j=1}^J \left( \sum_{i=1}^{k_j} W_{1ij}^* W_{2ij}^* - \frac{1}{k_j} \left( \sum_{i=1}^{k_j} W_{1ij}^* \sum_{i=1}^{k_j} W_{2ij}^* \right) \right) \right)^2 \end{aligned}$$

Finally,  $\Delta = \frac{\det(\Sigma) - \det(\Sigma^*)}{\det(\Sigma)}$  is the pseudo- $R^2$  criterion.

This quantity measures the global predictive accuracy of the PRS. As shown from the **Supplementary Material** and the

simulation study, the pseudo- $R^2$  is unit-less, ranges from zero to one, increases with the effect related with either the tail defect proportion or the dynamic of the occurrence of the event of interest.

## 2.5. Simulation-Based Studies

The objective of this section is to evaluate the behavior of the proposed index  $\Delta$  for different levels of tail defect, values of parameters  $\alpha$  and  $\beta$ , sample sizes  $n$  and percentages of censoring. We compare the values of  $\Delta$  to those of the pseudo- $R^2$  proposed by Nagelkerke ( $N$ ) based on a transformation of the partial likelihood ratio test (Nagelkerke, 1991), by O’Quigley and Flandre ( $OF$ ) based on Schoenfeld residuals (O’Quigley and Flandre, 1994), by Xu and O’Quigley ( $XO$ ) (Ronghui Xu and O’Quigley, 1999), by O’Quigley et al. ( $OXS$ ) (O’Quigley et al., 2005) based on explained randomness measures relying on a transformation of the Kullback-Leibler information gain and by Rouam et al. ( $RMB$ ) (Rouam et al., 2010) based on the robust score statistic. These indexes are implemented in the R packages “survAUC” (Potapov et al., 2015) (for  $N$ ,  $XO$ , and  $OXS$ ), “PHeval” (Chauvel, 2018) (for  $OF$ ), and “survival” (Therneau, 2015) (for  $RMB$ ).

For the sake of simplicity and without loss of generality, we evaluated the behavior of the test with only one stratum for most of the simulation scenarios. However, in order to check the behavior of the indice with strata, we performed an additional simulation scheme in a case with two strata.

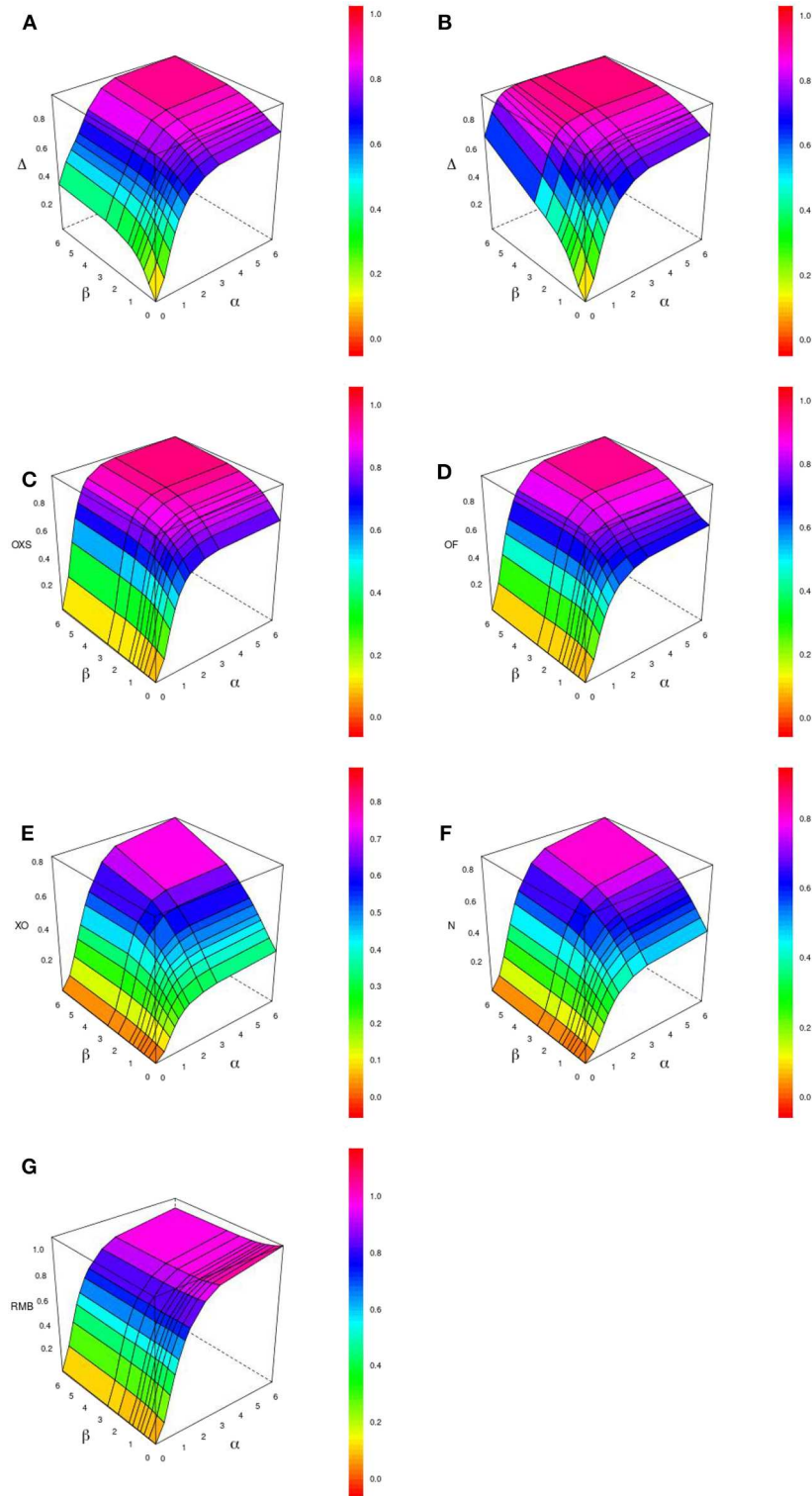
### 2.5.1. Simulation Scheme

Survival times were generated according to the two-part survival model presented in the previous section with baseline conditional cumulative hazard function  $\Lambda_0(s) = s$ . The baseline probability of being a non-susceptible was chosen such as  $\exp(-\theta)$  was equal to 0.3, 0.5, or 0.7. For each subject, we simulated a variable  $Z$  (its PRS) from a standard Normal distribution  $[N(0, 1)]$ .

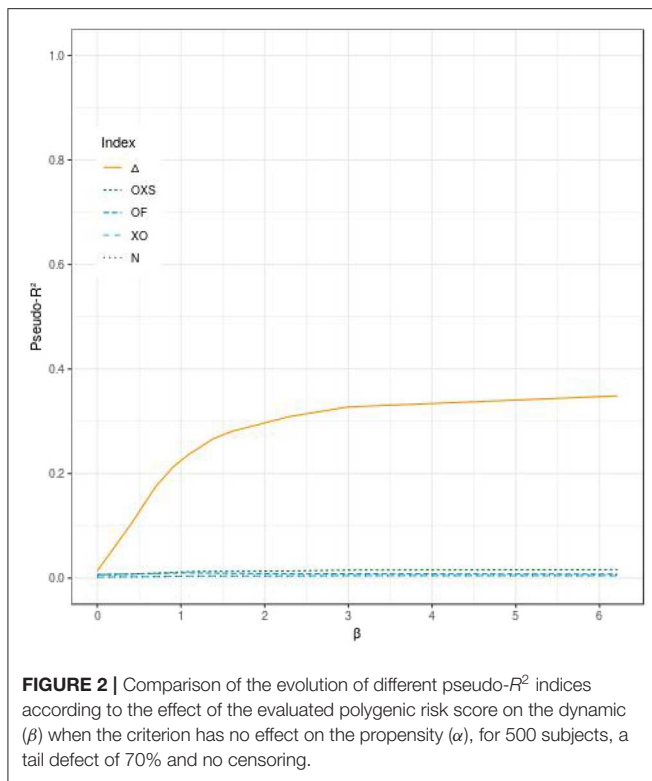
In order to see the behaviors of the different indices and in particular if they were able to attain values close to one, the following configurations were considered for the hazard ratio (HR) values  $e^\alpha$  and  $e^\beta$  (termed as “propensity effect” and “dynamic effect”, respectively): 1 (no effect), 1.5, 2, 2.5, 3, 4, 5, 10, 20, 50, or 500. These values explore a large range of effects from small [ $\log(1.5) \approx 0.41$ ] to huge [ $\log(500) \approx 6.21$ ] for  $\alpha$  and  $\beta$ . The number of subjects was 500.

For investigating the robustness of the proposed pseudo- $R^2$  to model misspecification, we performed simulations with survival times generated according to an improper Gompertz distribution such as  $S(s|Z = z) = \exp(-\theta e^{\alpha z} (1 - \exp(-se^{\beta z}))$ .

For investigating the robustness of the indice to the distribution of the covariate, we simulated a variable  $Z$  from a Student distribution with ten degrees of freedom. The effect of censoring was investigated by generating independently censoring times from a uniform distribution over  $[0, u]$ . Values for  $u$  were computed from the chosen percentage of censoring and from the parameters of the considered distributions. The percentage of censoring refers to the percentage of censored



**FIGURE 1** | Display of the behavior of different pseudo- $R^2$  indices when the effects of the evaluated criterion on the propensity ( $\alpha$ ) and on the dynamic ( $\beta$ ) vary. The graphic shows  $\Delta$  with 0% **(A)** and 20% **(B)** censoring and other indexes with 0% censoring : OXS **(C)**, OF **(D)**, XO **(E)**, N **(F)**, and RMB **(G)**, for 500 subjects and a tail defect of 70%.



observations without the fraction of non-susceptible subjects. Here, 20% censoring were considered.

For all these simulations, values  $e^\alpha$  and  $e^\beta$  were :1, 1.25, 1.5 and 1, 1.5, 2, respectively. The number of subjects was 500.

For investigating the robustness of the proposed pseudo- $R^2$  to the number of subjects, we performed additional simulations with 200, 500, and 1,000 subjects. We also performed additional simulations where we generated a stratification variable from a Bernoulli distribution with parameter 0.5.

For each configuration, 1,000 replications were performed.

### 3. RESULTS

#### 3.1. Simulation Results

**Figure 1** displays the behavior of  $\Delta$ ,  $OF$ ,  $OXS$ ,  $XO$ ,  $N$ , and  $RMB$  according to  $\alpha$  and  $\beta$  for a tail defect of 70% with 500 subjects and no censoring. Our simulations showed that the value of  $\Delta$  increases with the value of the strength of the PRS' effect (through the hazards ratio  $e^\alpha$  and  $e^\beta$ ). In our simulation study,  $\Delta$  ranges from 0 when both  $\alpha$  and  $\beta$  are null, to near 1 for very large effects of  $\alpha$  and  $\beta$ . In the range of effects presented in this paper, the maximum value reached by  $\Delta$  is 0.96 when both  $\alpha$  and  $\beta$  take value of 6.21 (corresponding to a HR of 500) (see **Figure 1A**). We can see that  $\Delta$  is able to quantify an effect on the dynamic ( $\beta$ ) when there is no propensity effect ( $\alpha$ ) related to the PRS. For example, when  $\alpha$  is null and  $\beta = 0.92$  (corresponding to a HR of  $e^\beta = 2.5$ ) we have  $\Delta = 0.52$ . The other studied indices are not able to quantify a dynamic effect when there is no propensity effect (see **Figures 1A,C-F, 2**).

When the propensity effect is not null, all indices, including ours, are able to quantify the mixture of both effects. The different configurations show that  $\Delta$ ,  $OF$ , and  $OXS$  lead to higher values than  $XO$  and  $N$ . **Figure 3** displays the values of the indices according to the propensity and dynamic effects, when both the parameters  $\alpha$  and  $\beta$  have the same value.  $\Delta$  has the highest values when  $\alpha = \beta \leq 0.91$  ( $e^\alpha = e^\beta \leq 2.5$ ) and ranks in third place for higher values of  $\alpha$  and  $\beta$ .

**Figure 4** shows that  $\Delta$  quantifies both the propensity effect and the dynamic effect. However, the reported predictive accuracy is not symmetrical and is higher for the propensity effect ( $\alpha$ ) as compared to the dynamic effect ( $\beta$ ). For example, when  $\alpha$  is null and  $\beta = 0.92$  (HR of 2.5),  $\Delta = 0.21$ , whereas when  $\alpha = 0.92$  and  $\beta$  is null, then  $\Delta = 0.42$  (see **Figure 4**).

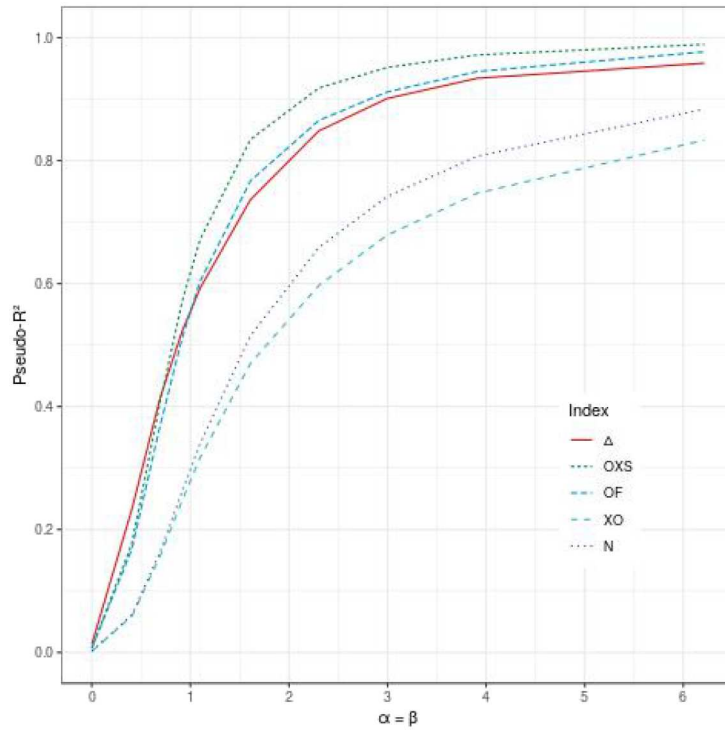
**Table 1** displays the results obtained with the pseudo- $R^2$ s for uncensored cases with various tail defects. These results show that the proposed pseudo- $R^2$  is the only one able to quantify the dynamic effect. As an example, for a tail defect of 30% with  $e^\alpha = 1.25$  and  $e^\beta = 2.50$ , its value is of 37% whereas the highest value for the other indices is 16% (with  $OXS$ ). The same behavior of the proposed pseudo- $R^2$  is observed for the different values of the tail defect.

**Table 2** displays the results obtained for the proposed pseudo- $R^2$  under various tail defects and simulation schemes: (i) Two-part survival model with no censoring and Normally distributed explanatory variables, (ii) Two-part survival model with 20% censoring and Normally distributed explanatory variables, (iii) Two-part survival model with no censoring and Student distributed explanatory variables, (iv) Improper Gompertz model with no censoring and Normally distributed explanatory variables.

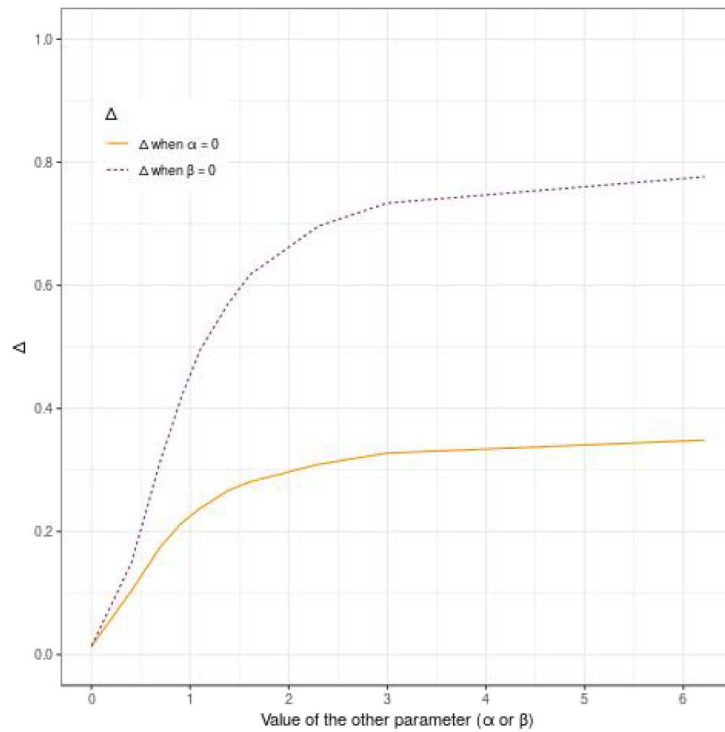
In case of 20% censoring, simulation results show that the mean and standard error values of the proposed indice are slightly increased. When looking to the simulation results under an improper Gompertz model, simulation results show that the proposed indice is only slightly affected by model misspecification. For a Student distribution, we can see that the estimated mean values of the proposed indice are very close to the Normal distribution. **Table 3** shows that the estimated mean and standard error values of the proposed indice are slightly increased when the number of subjects decreases. It is worth noting that the slight bias observed for the mean values with censored or reduced samples is linked to the fact that our pseudo- $R^2$  uses the estimate of the tail defect. In these situations, the dispersion of this nuisance parameter increases.

As expected, results obtained for the proposed pseudo- $R^2$ s with and without strata are the same (**Table 4**).

It should be noted that comparison of the values of the pseudo- $R^2$  across different tail defect values are difficult to interpret and can be misleading since the dynamic/propensity effects should be interpreted conditionally upon the defective survival distributions that are not normalized to one but to different values according to the defect. This is also the case for the other indices that increase when the proportion of susceptible subjects increase. For example, with  $e^\alpha = 1.5$  and  $e^\beta = 2.5$ , values of the  $OXS$  are 29, 25, and 20% for a tail defect of 30, 50, and 70%, respectively.



**FIGURE 3 |** Comparison of the evolution of different pseudo- $R^2$  indices according to the effect of the evaluated polygenic risk score on the dynamic ( $\beta$ ) and the propensity ( $\alpha$ ), when  $\alpha = \beta$ , for 500 subjects, a tail defect of 70% and no censoring.



**FIGURE 4 |** Display of the behavior of the pseudo- $R^2$   $\Delta$  according to the effect of the evaluated polygenic risk score on the dynamic ( $\beta$ ) or the propensity ( $\alpha$ ), when  $\alpha$  or  $\beta$  is null, for 500 subjects, a tail defect of 70% and no censoring.



**TABLE 1** | Comparison of the different pseudo- $R^2$  indices in simulations.

Index	$e^\alpha$	1			1.25			1.5		
	$e^\beta$	1	2	2.5	1	2	2.5	1	2	2.5
OXS	Mean	0.00	0.03	0.05	0.04	0.14	0.16	0.11	0.25	0.29
	SD	0.01	0.02	0.02	0.02	0.04	0.04	0.03	0.05	0.04
XO	Mean	0.00	0.02	0.03	0.03	0.10	0.12	0.07	0.18	0.21
	SD	0.00	0.01	0.02	0.01	0.03	0.03	0.02	0.03	0.03
N	Mean	0.00	0.02	0.03	0.03	0.10	0.12	0.07	0.18	0.21
	SD	0.00	0.01	0.02	0.01	0.03	0.03	0.02	0.03	0.03
OF	Mean	0.00	0.02	0.03	0.04	0.12	0.14	0.10	0.24	0.26
	SD	0.01	0.02	0.02	0.02	0.04	0.04	0.03	0.04	0.04
RMB	Mean	0.00	0.03	0.04	0.04	0.12	0.14	0.11	0.23	0.25
	SD	0.01	0.02	0.02	0.02	0.03	0.04	0.03	0.04	0.04
$\Delta$	Mean	0.01	0.21	0.27	0.06	0.31	0.37	0.15	0.39	0.45
	SD	0.01	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.04

**(A) Tail defect of 30%**

Index	$e^\alpha$	1			1.25			1.5		
	$e^\beta$	1	2	2.5	1	2	2.5	1	2	2.5
OXS	Mean	0.00	0.02	0.02	0.05	0.10	0.11	0.13	0.22	0.24
	SD	0.01	0.01	0.02	0.03	0.04	0.04	0.04	0.05	0.05
XO	Mean	0.00	0.01	0.01	0.02	0.05	0.06	0.07	0.12	0.13
	SD	0.00	0.01	0.01	0.01	0.02	0.02	0.02	0.03	0.03
n	Mean	0.00	0.01	0.01	0.02	0.05	0.06	0.07	0.12	0.13
	SD	0.00	0.01	0.01	0.01	0.02	0.02	0.02	0.03	0.03
OF	Mean	0.00	0.01	0.01	0.05	0.09	0.10	0.13	0.21	0.22
	SD	0.01	0.01	0.01	0.03	0.04	0.04	0.04	0.05	0.05
RMB	Mean	0.00	0.01	0.02	0.05	0.09	0.10	0.14	0.20	0.21
	SD	0.01	0.01	0.02	0.03	0.03	0.04	0.04	0.05	0.05
$\Delta$	Mean	0.01	0.18	0.22	0.06	0.26	0.31	0.15	0.33	0.38
	SD	0.01	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04

**(B) Tail defect of 50%**

Index	$e^\alpha$	1			1.25			1.5		
	$e^\beta$	1	2	2.5	1	2	2.5	1	2	2.5
OXS	Mean	0.01	0.01	0.01	0.05	0.08	0.08	0.14	0.19	0.20
	SD	0.01	0.01	0.01	0.03	0.04	0.04	0.05	0.06	0.06
XO	Mean	0.00	0.00	0.00	0.02	0.02	0.03	0.05	0.07	0.07
	SD	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.02	0.02
N	Mean	0.00	0.00	0.00	0.02	0.03	0.03	0.05	0.07	0.07
	SD	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.02	0.02
OF	Mean	0.01	0.01	0.01	0.05	0.07	0.08	0.14	0.18	0.18
	SD	0.01	0.01	0.01	0.03	0.04	0.04	0.05	0.06	0.06
RMB	Mean	0.01	0.01	0.01	0.05	0.07	0.08	0.14	0.17	0.18
	SD	0.01	0.01	0.01	0.03	0.04	0.04	0.05	0.05	0.05
$\Delta$	Mean	0.01	0.17	0.21	0.06	0.24	0.28	0.15	0.31	0.35
	SD	0.01	0.04	0.04	0.03	0.05	0.04	0.05	0.05	0.04

**(C) Tail defect of 70%.**

OXS, pseudo- $R^2$  index from O'Quigley et al. (2005); N, pseudo- $R^2$  index from Nagelkerke (1991); XO, pseudo- $R^2$  index from Ronghui Xu and O' quigley (1999); OF, pseudo- $R^2$  index from O'Quigley and Flandre (1994); RMB, pseudo- $R^2$  index from Rouam et al. (2010);  $\Delta$ , the pseudo- $R^2$  index we propose in this work;  $e^\alpha$ , propensity effect;  $e^\beta$ , dynamic effect; SD, standard deviation. Simulations were performed with 500 subjects.

**TABLE 2** |  $\Delta$  pseudo- $R^2$  indices in simulations with various simulation designs.

Design	$e^\alpha$	1			1.25			1.5		
	$e^\beta$	1	2	2.5	1	2	2.5	1	2	2.5
(i)	Mean	0.01	0.21	0.27	0.06	0.31	0.37	0.15	0.39	0.45
	SD	0.01	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.04
(ii)	Mean	0.01	0.22	0.29	0.05	0.31	0.39	0.12	0.40	0.46
	SD	0.01	0.05	0.04	0.03	0.05	0.05	0.04	0.05	0.05
(iii)	Mean	0.01	0.21	0.27	0.08	0.32	0.38	0.17	0.41	0.46
	SD	0.01	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.04
(iv)	Mean	0.01	0.19	0.25	0.08	0.31	0.37	0.19	0.41	0.46
	SD	0.01	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.04

**(A) Tail defect of 30%**

Design	$e^\alpha$	1			1.25			1.5		
	$e^\beta$	1	2	2.5	1	2	2.5	1	2	2.5
(i)	Mean	0.01	0.18	0.22	0.06	0.26	0.31	0.15	0.33	0.38
	SD	0.01	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04
(ii)	Mean	0.01	0.18	0.25	0.05	0.27	0.33	0.13	0.35	0.40
	SD	0.01	0.04	0.04	0.03	0.05	0.05	0.04	0.05	0.05
(iii)	Mean	0.01	0.18	0.22	0.07	0.27	0.31	0.17	0.35	0.39
	SD	0.01	0.03	0.04	0.03	0.04	0.04	0.04	0.04	0.04
(iv)	Mean	0.01	0.17	0.21	0.06	0.26	0.31	0.16	0.34	0.38
	SD	0.01	0.03	0.04	0.03	0.04	0.04	0.05	0.04	0.04

**(B) Tail defect of 50%**

Design	$e^\alpha$	1			1.25			1.5		
	$e^\beta$	1	2	2.5	1	2	2.5	1	2	2.5
(i)	Mean	0.01	0.17	0.21	0.06	0.24	0.28	0.15	0.31	0.35
	SD	0.01	0.04	0.04	0.03	0.05	0.04	0.05	0.05	0.04
(ii)	Mean	0.02	0.17	0.23	0.06	0.24	0.30	0.14	0.33	0.37
	SD	0.02	0.05	0.05	0.04	0.06	0.05	0.06	0.06	0.05
(iii)	Mean	0.01	0.17	0.21	0.07	0.25	0.29	0.17	0.32	0.35
	SD	0.01	0.04	0.04	0.04	0.04	0.04	0.05	0.04	0.04
(iv)	Mean	0.01	0.17	0.21	0.06	0.24	0.29	0.15	0.31	0.35
	SD	0.01	0.04	0.04	0.04	0.05	0.04	0.05	0.05	0.04

**(C) Tail defect of 70%**

$e^\alpha$ , propensity effect;  $e^\beta$ , dynamic effect; SD, standard deviation. Simulations were performed with 500 subjects. Detail of the designs displayed in the table : (i) two-part survival model with no censoring and Normally distributed explanatory variables; (ii) Two-part survival model with 20% censoring and Normally distributed explanatory variables (iii) Two-part survival model with no censoring and Student distributed explanatory variables; (iv) Improper Gompertz model with no censoring and Normally distributed explanatory variables.

**TABLE 3** |  $\Delta$  pseudo- $R^2$  indices in simulations with various number of subjects.

Subjects	$e^\alpha$	1			1.25			1.5		
	$e^\beta$	1	2	2.5	1	2	2.5	1	2	2.5
200	Mean	0.02	0.23	0.29	0.07	0.32	0.38	0.16	0.39	0.45
	SD	0.02	0.06	0.06	0.04	0.07	0.07	0.06	0.07	0.06
500	Mean	0.01	0.21	0.27	0.06	0.31	0.37	0.15	0.39	0.45
	SD	0.01	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.04
1,000	Mean	0.00	0.20	0.26	0.06	0.30	0.37	0.15	0.39	0.45
	SD	0.00	0.03	0.03	0.02	0.03	0.03	0.03	0.03	0.03

**(A) Tail defect of 30%**

Subjects	$e^\alpha$	1			1.25			1.5		
	$e^\beta$	1	2	2.5	1	2	2.5	1	2	2.5
200	Mean	0.02	0.20	0.25	0.07	0.27	0.32	0.15	0.33	0.39
	SD	0.02	0.05	0.06	0.05	0.06	0.06	0.06	0.06	0.06
500	Mean	0.01	0.18	0.22	0.06	0.26	0.31	0.15	0.33	0.38
	SD	0.01	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04
1,000	Mean	0.00	0.17	0.21	0.05	0.25	0.30	0.15	0.33	0.38
	SD	0.00	0.02	0.03	0.02	0.03	0.03	0.03	0.03	0.03

**(B) Tail defect of 50%**

Subjects	$e^\alpha$	1			1.25			1.5		
	$e^\beta$	1	2	2.5	1	2	2.5	1	2	2.5
200	Mean	0.04	0.20	0.24	0.08	0.25	0.30	0.16	0.32	0.36
	SD	0.03	0.06	0.06	0.06	0.08	0.07	0.08	0.08	0.07
500	Mean	0.01	0.17	0.21	0.06	0.24	0.28	0.15	0.31	0.35
	SD	0.01	0.04	0.04	0.03	0.05	0.04	0.05	0.05	0.04
1,000	Mean	0.01	0.16	0.20	0.05	0.23	0.28	0.14	0.30	0.34
	SD	0.01	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.03

**(C) Tail defect of 70%**

$e^\alpha$  : propensity effect;  $e^\beta$  : dynamic effect; SD : standard deviation.

**TABLE 4** |  $\Delta$  pseudo- $R^2$  index in simulations with strata.

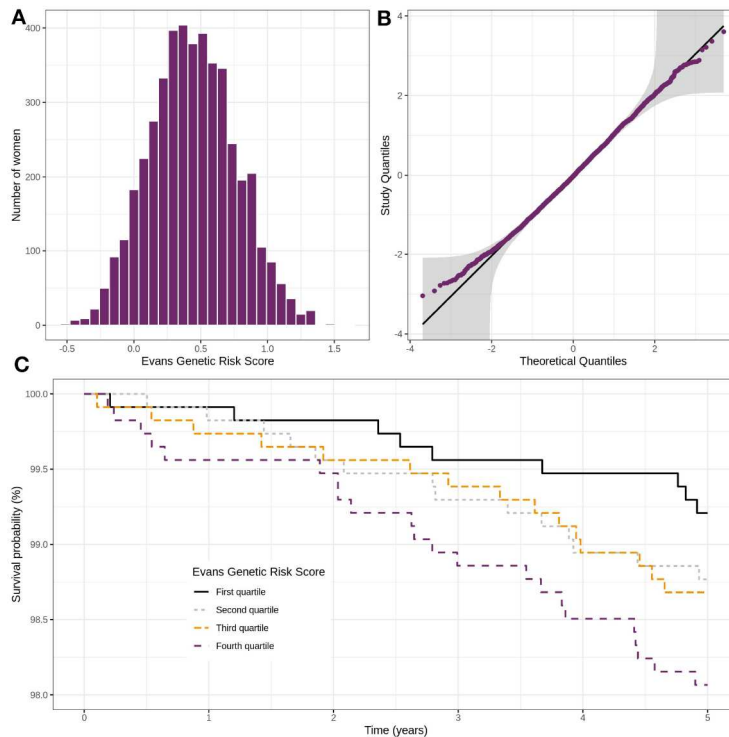
Tail defect	$e^\alpha$	1			1.25			1.5		
	$e^\beta$	1	2	2.5	1	2	2.5	1	2	2.5
0.3	Mean	0.01	0.22	0.28	0.06	0.31	0.37	0.15	0.39	0.45
	SD	0.01	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.04
0.5	Mean	0.01	0.19	0.23	0.06	0.26	0.31	0.15	0.33	0.38
	SD	0.01	0.03	0.04	0.03	0.04	0.04	0.04	0.04	0.04
0.7	Mean	0.01	0.18	0.23	0.06	0.25	0.29	0.15	0.30	0.35
	SD	0.01	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.04

$e^\alpha$ , propensity effect;  $e^\beta$ , dynamic effect; SD, standard deviation.

### 3.2. CARTaGENE Results

Among the 4,554 women selected for the analysis, 60 (1.32%) had breast cancer during the 5 years of follow-up. The PRS' mean

(Evan's score) was higher among participants with a diagnosed breast cancer within the 5 years (0.57) than for those free of event at 5 years (0.44).



**FIGURE 5** | Distribution of Evans' polygenic risk score in the CARTaGENE cohort ( $n = 4,554$ ) **(A)**, control women QQ-plot with confidence bands based on an inversion of the Kolmogorov–Smirnov test **(B)** and incidence of breast cancer in the cohort according to the Evans' score **(C)**.

**Figure 5** displays the distribution of the PRS in the cohort, the normal quantile-quantile plot and the Kaplan-Meier estimate of the probability of not experiencing breast cancer for the four groups based on the PRS quartiles. As seen from **Figures 5A,B**, the distribution of the PRS can be considered close to the Normal distribution. From **Figure 5C**, we can see that higher the PRS, higher the risk of breast cancer is.

**Table 5** shows the results for  $\Delta$  and the five pseudo- $R^2$  indices. For these latter, pseudo- $R^2$  measures are computed upon an entry-age-stratified age-scaled Cox survival model.  $\Delta$  was the highest pseudo- $R^2$  (17.8%), while *OF*, *RMB*, and *OXS* were equal to 12.0, 12.1, and 14.0%, respectively. The indices *XO* and *N* had values of 0.21 and 1.01%, respectively.

## DISCUSSION

The evaluation of the PRS for clinical prediction has received a lot of attention these last years. For binary or time-to-event traits, prediction accuracy of PRS is usually assessed using likelihood-based pseudo- $R^2$  criteria that can be complicated to compute for complex models. In this work, we have proposed a novel pseudo- $R^2$  test for assessing the accuracy prediction over a time period of PRS for both the probability of occurrence of the event of interest and the dynamic. This criterion is easy to compute since it avoids maximizing Cox's partial likelihood under the alternative.

As seen from the simulation study, our pseudo- $R^2$  showed good performances as compared to classical indices based on the Cox proportional hazards model. As expected, it showed

**TABLE 5** | Pseudo- $R^2$  index values for Evans genetic risk score in the CARTaGENE cohort ( $n = 4,554$ ).

	Pseudo- $R^2$ (%)
OXS	14.1
N	1.01
XO	0.21
OF	12.0
RMB	12.1
$\Delta$	17.8

*OXS*, pseudo- $R^2$  index from O'Quigley et al. (2005); *N*, pseudo- $R^2$  index from Nagelkerke (1991); *XO*, pseudo- $R^2$  index from Ronghui Xu and O' quigley (1999); *OF*, pseudo- $R^2$  index from O'Quigley and Flandre (1994); *RMB*, pseudo- $R^2$  index from Rouam et al. (2010);  $\Delta$ , the pseudo- $R^2$  index we propose in this work.

very good performances when there is a dynamic effect related to the PRS. Moreover, when the main effect of the PRS is on the dynamic of the event among the susceptible subjects, the proposed pseudo- $R^2$  is the only one able to quantify this effect. Our simulations have shown the good behaviors of the indices proposed by O'Quigley and Flandre (based on Schoenfeld residuals) (O'Quigley and Flandre, 1994), by O'Quigley et al. (based on explained randomness measures) (O'Quigley et al., 2005), and by Rouam et al. (2010). In contrast, the coefficient proposed by Nagelkerke (1991) that is frequently used in the literature showed poor results. Simulations results have shown that the proposed indice is only barely affected by model

misspecification and covariate distribution. It slightly increases with reduced sample size or censoring. Based on our simulation study, when a dynamical effect related to the PRS is expected, we recommend using our proposed pseudo- $R^2$ . In other cases, our index can be used together with the *OF* and *OXS* indices since they show good performances. In contrast, we do not recommend the use of the Nagelkerke index for quantifying predictive accuracy of PRS for time to event outcomes.

Notwithstanding the good performance of the proposed criteria, it has some shortcomings that should be mentioned. We should first keep in mind that pseudo- $R^2$ s are always model-based criteria. In our case, our pseudo- $R^2$  measures the proportion of variability in the outcome that is explained by the PRS under the assumption that the PRS may be linked to the dynamic and/or the propensity of the occurrence of the event. In this work, it relies on a two-component survival model which is potentially prone to misspecification. Thus, such underlying assumption of a mixture model should be discussed before considering the use of the proposed pseudo- $R^2$ . We do not claim that our model represents the reality but that it is a useful approximation of what we suppose the effects (propensity/dynamic) are. Moreover, it is also worth noting that pseudo- $R^2$ s and even the classical  $R^2$  are not robust to outlier observations and non-normal distributions. Thus, we should warn practitioners to be cautious and examine the PRS distribution. Here, we have considered pseudo- $R^2$  criteria since our main interest was to focus on the percentage of variation in the outcome explained by the PRS. Time-dependent ROC curves (with incident cases/dynamic controls; Heagerty and Zheng, 2005) could also have been considered if the main objective was to evaluate the prognostic potential of the PRS by focusing on the correct classification rates. Moreover, it should be noted that our pseudo- $R^2$  is not designed for more complex situations such as those with marker-dependent censoring. We plan to do further work in this direction.

This novel pseudo- $R^2$  was used to evaluate the 5-years predictive accuracy of the PRS proposed by Evans et al. (2017) for breast cancer occurrence on the CARTaGENE cohort. In this study, we reported a pseudo- $R^2$  of 17.8% for our novel index which is higher than the values reported by *OF* and *OXS* indices. Based on the results from our simulation study, we hypothesize that the PRS has both a propensity effect and a dynamical effect. This is not surprising since the selected SNP are located within genes that encode for proteins that are involved in important processes such as cell growth and division. It is worth noting that for this analysis we have considered an age-dependent model with four strata. Here, our pseudo- $R^2$  provides a global predictive measure that average all differences for both the propensity and dynamic effects taking into account the stratum of age at entry. Other strategies can be considered and easily implemented. We should keep in mind that the majority of PRS were developed and evaluated from case-control designs which raises some issues about misspecification when applying these results for time-to-event prediction (Lambert et al., 2019). With large prospective cohorts, we may expect to see in a close future more published polygenic hazard scores.

Finally, we think that the proposed novel pseudo- $R^2$ , which is easy to implement with standard softwares, is worth being used to evaluate PRS for predicting incident events in cohort studies.

## DATA AVAILABILITY STATEMENT

The data analyzed in this study were collected in the context of the CARTaGENE cohort. The data are available upon official request and ethical approval. The dataset cannot be made publicly available but can be obtained by interested researchers upon request to the CARTaGENE team at CHU Sainte-Justine. The requests can be made using the CARTaGENE online application at [sdas.cartagene.qc.ca](https://sdas.cartagene.qc.ca), by contacting the CARTaGENE team at [access@cartagene.qc.ca](mailto:access@cartagene.qc.ca) or by calling 514-345-2156.

## ETHICS STATEMENT

CARTaGENE has obtained ethics approval by the CHU Sainte-Justine under the reference: MP-21-2011-345, 3297. The latest annual ethics renewal was granted on September 13, 2019.

## AUTHOR CONTRIBUTIONS

JD and PB developed the original statistic. PB coordinated the project and is JD's Ph.D. thesis advisor. JD, RJ, and PB analyzed the data. TD, YP, CL, and NN participated in the collection of the data. JD, RJ, TD, YP, CL, NN, and PB participated in writing the original draft. All authors read and approved the final manuscript.

## FUNDING

The research of the first author was supported by a grant from University Paris-Saclay (France). The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## ACKNOWLEDGMENTS

This research work has been motivated by questions raised by the CARTaGENE cohort at the CHU Sainte-Justine. The authors would like to thank all the CARTaGENE participants for their generous investments in health research. The authors would also like to thank the Régie de l'assurance maladie du Québec (RAMQ) and the Commission d'accès à l'information (CAI) for their support in obtaining the data relevant to the study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00408/full#supplementary-material>

## REFERENCES

- Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P., and Zondervan, K. T. (2010). Data quality control in genetic case-control association studies. *Nat. Protoc.* 5, 1564–1573. doi: 10.1038/nprot.2010.116
- Awadalla, P., Boileau, C., Payette, Y., Idaghmour, Y., Goulet, J.-P., Knoppers, B., et al. (2013). Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. *Int. J. Epidemiol.* 42, 1285–1299. doi: 10.1093/ije/dys160
- Buniello, A., MacArthur, J. A., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47, D1005–D1012. doi: 10.1093/nar/gky1120
- Chauvel, C. (2018). *PHeval: Evaluation of the Proportional Hazards Assumption with a Standardized Score Process*. CRAN. Available online at: <https://CRAN.Rproject.org/package=PHeval>
- Cox, D. R. (1972). Regression models and life-tables. *J. R. Stat. Soc. Ser. B* 34, 187–220. doi: 10.1111/j.2517-6161.1972.tb00899.x
- Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., et al. (2016). Next-generation genotype imputation service and methods. *Nat. Genet.* 48, 1284–1287. doi: 10.1038/ng.3656
- Evans, D. G., Brentnall, A., Byers, H., Harkness, E., Stavrinou, P., Howell, A., et al. (2017). The impact of a panel of 18 SNPs on breast cancer risk in women attending a UK familial screening clinic: a case-control study. *J. Med. Genet.* 54, 111–113. doi: 10.1136/jmedgenet-2016-104125
- Flandre, P., Deutsch, R., and O'Quigley, J. (2017). Accuracy of predictive ability measures for survival models. *Stat. Med.* 36, 3171–3180. doi: 10.1002/sim.7342
- Fleming, T., and Harrington, D. (2005). *Counting Processes and Survival Analysis*. New York, NY: John Wiley & Sons, Inc.
- Heagerty, P. J., and Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics* 61, 92–105. doi: 10.1111/j.0006-341X.2005.030814.x
- Hu, B., Shao, J., and Palta, M. (2006). Pseudo-R<sup>2</sup> in logistic regression model. *Stat. Sin.* 16, 847–860. Available online at: [www.jstor.org/stable/24307577](http://www.jstor.org/stable/24307577)
- International Multiple Sclerosis Genetics Consortium, I., Bush, W., Sawcer, S., de Jager, P., Oksenberg, J., McCauley, J., et al. (2010). Evidence for polygenic susceptibility to multiple sclerosis - the shape of things to come. *Nat. Genet.* 42, 621–625. doi: 10.1016/j.ajhg.2010.02.027
- Khera, A., Chaffin, M., Aragam, K., Haas, M., Roselli, C., Choi, S., et al. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* 50, 1219–1224. doi: 10.1038/s41588-018-0183-z
- Lachin, J. M. (2011). *Biostatistical Methods: The Assessment of Relative Risks*. Hoboken, NJ: John Wiley & Sons Ltd.
- Lambert, S. A., Abraham, G., and Inouye, M. (2019). Towards clinical utility of polygenic risk scores. *Hum. Mol. Genet.* 28, R133–R142. doi: 10.1093/hmg/ddz187
- Lin, D. Y., and Wei, L. J. (1989). The robust inference for the cox proportional hazards model. *J. Am. Stat. Assoc.* 84, 1074–1078. doi: 10.1080/01621459.1989.10478874
- Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., A Reshef, Y., K Finucane, H., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* 48, 1443–1448. doi: 10.1038/ng.3679
- Machiela, M., Chen, C., Chen, C., Chanoock, S., Hunter, D., and Kraft, P. (2011). Evaluation of polygenic risk scores for predicting breast and prostate cancer risk. *Genet. Epidemiol.* 35, 506–514. doi: 10.1002/gepi.20600
- Maller, R. A., and Zhou, X. (1996). *Survival Analysis With Long-Term Survivors*. Wiley series in probability and statistics. New York, NY; Chichester: Wiley.
- McCarthy, M., Abecasis, G., Cardon, L., Goldstein, D., Little, J., Ioannidis, J., et al. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9, 356–369. doi: 10.1038/nrg2344
- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika* 78, 691–692. doi: 10.1093/biomet/78.3.691
- O'Quigley, J. (2008). *Proportional Hazards Regression*. New York, NY: Springer Science+Business Media, LLC.
- O'Quigley, J., and Flandre, P. (1994). Predictive capability of proportional hazards regression. *Proc. Natl. Acad. Sci. U.S.A.* 91, 2310–2314. doi: 10.1073/pnas.91.6.2310
- O'Quigley, J., Xu, R., and Stare, J. (2005). Explained randomness in proportional hazards models. *Stat. Med.* 24, 479–489. doi: 10.1002/sim.1946
- Perdry, H., and Dandine-Roulland, C. (2019). *Gaston: Genetic Data Handling (QC, GRM, LD, PCA) & Linear Mixed Models*. CRAN. Available online at: <https://CRAN.Rproject.org/package=gaston>
- Potapov, S., Adler, W., and Schmid, M. (2015). *survAUC: Estimators of Prediction Accuracy for Time-to-Event Data*. CRAN. Available online at: <https://CRAN.Rproject.org/package=survAUC>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Ronghui X., and O' quigley, J. (1999). A R<sup>2</sup> type measure of dependence for proportional hazards models. *J. Nonparamet. Stat.* 12, 83–107. doi: 10.1080/10485259908832799
- Rouam, S., Moreau, T., and Broët, P. (2010). Identifying common prognostic factors in genomic cancer studies: A novel index for censored outcomes. *BMC Bioinformatics* 11:150. doi: 10.1186/1471-2105-11-150
- The International HapMap 3 Consortium (2010). Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58. doi: 10.1038/nature09298
- Théberge, I., Institut national de santé publique du Québec, and Direction systèmes de soins et services (2003). *Validation de Stratégies pour Obtenir le Taux de Détection du Cancer...* Direction des systèmes de soins et services. Institut National de Santé Publique.
- Therneau, T. M. (2015). *A Package for Survival Analysis in S*. version 2.38. Available online at: <https://CRAN.R-project.org/package=survival>
- Torkamani, A., Wineinger, N., and Topol, E. (2018). The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* 19, 581–590. doi: 10.1038/s41576-018-0018-x

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Duhazé, Jantzen, Payette, De Malliard, Labbé, Noisel and Broët. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.