Wiley StatsRef-Statistics Reference Online:

Dynamic Predictions

Cécile Proust-Lima [*,a,b], and Paul Blanche [c]

November 2015

[a] INSERM, U1219, Bordeaux Population Health Research Center, F-33076 Bordeaux, France
[b] Université de Bordeaux, ISPED, F-33076 Bordeaux, France
[c] LMBA, Université de Bretagne sud, Vannes, France

[*] correspondence to:
Cécile Proust-Lima, Inserm U1219, ISPED, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France.
E-mail: cecile.proust-lima@inserm.fr

Abstract:
Individualized risk prediction has become central in medicine, especially in the monitoring, screening and management of chronic diseases. Dynamic predictions consist of predictions that are based on information repeatedly collected over time and that can be updated (dynamically) as soon as new information becomes available. The so-called dynamic information typically includes biomarker measures and exposure records. The aim of this paper is to give a short overview about how dynamic predictions can be defined and computed, and how their predictive accuracy can be evaluated. References are provided to guide further reading about popular methods and examples.

Keywords:
biomarkers, joint model, landmark approach, longitudinal data, risk prediction, survival data

1 Introduction

Predicting the risk of an event based on individual information has become central in medicine, especially in the monitoring, screening and management of chronic diseases.

After a diagnosis of cancer for example, patients often seek answers to questions such as *"what is my risk of dying?", "what is my risk of experiencing a recurrence?", "what is my prognosis?"*. Providing them with risk estimates may help them, as well as their doctors, to better manage their condition. Depending on the risk that their patients have to experience progressions, doctors may adapt medical decisions, for instance by changing the treatment and the frequency of the follow-up visits. In other words, the treatment and monitoring of the disease can be personalized. Even prior to diagnosis, individualized risk prediction can help personalize screening strategies and/or guide the initiation of treatments to prevent adverse outcomes. Risk prediction can also facilitate earlier diagnoses, which is valuable because most of the time, the earlier the diagnosis the better the chance of treatment success. Finally, at the research level, individualized risk prediction may further help design preventive clinical trials by providing scores which facilitate the inclusion of subjects with specific risk profiles.

Individual risk prediction always brings the notion of time into play. Time is involved in two ways:

- A patient can ask what is his/her risk of dying in the coming t= 2, 3 or 5 years. This prediction window t is called the **horizon** of prediction.
- A patient can ask what is his/her risk of dying in the t years given he/she is alive after s= 6 months, 1 year or 4 years. This time s from which the prediction is made is called the **landmark time** of prediction.

In the setting in which some landmark times $s > 0$ are of interest, predictions are called **dynamic predictions**. The term "dynamic" emphasizes the fact that the computation of predictions can be updated at any landmark time after the start of the follow-up. Indeed the extra information collected during the time period that ranges from the beginning of the follow-up to the landmark time s, sometimes called the **dynamic information**, can be accounted for in the computation. The dynamic information can include the knowledge that the patient has survived up to the landmark time s only, as in the above example, but typically further includes information from relevant biomarkers, exposure records and/or treatments regimes recorded up to time s.

2 How are dynamic predictions defined?

The dynamic prediction framework is illustrated in Figure 1, within the context of kidney transplantation. The probability of the event of interest -here the graft failure- is computed at the landmark time s=4 years and accounts for the records from previous measures of a biomarker -here serum creatinine. The dynamic prediction is computed for a 5-year horizon and gives a 36% risk of experiencing a graft failure within these 5 years.

Formally, let T_i be the time to an event of interest and $\mathcal{H}_i(s)$ be the information collected for subject i ($i = 1, ..., N$) by time s. It usually includes information from baseline covariates X_i as well as repeated information on a set of K longitudinal covariates $Y_{ik}(t_{kij})$ collected at

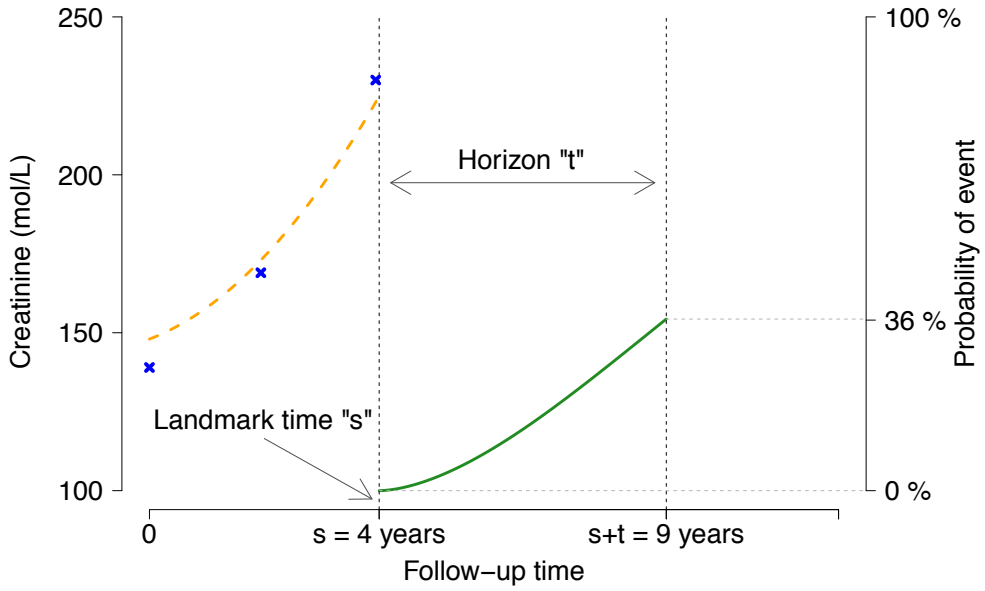

Figure 1: Framework of dynamic predictions in the context of kidney transplantation. The cumulative conditional probability of graft failure (plain line) is computed from the landmark time s=4 years for all t ranging up o a maximum horizon of 5 years. It is computed conditional on the 3 repeated measurements (crosses) of the serum creatinine previously collected prior to the landmark time s. The creatinine measurements enable the modeling of the subject-specific creatinine trajectory (dashed line), which provides valuable information about the subject-specific risk of graft failure.

times t_{kij} ($k = 1, ..., K$, $j = 1, ..., n_{ki}$). With these notations the individual dynamic prediction of experiencing the event of interest between times s and $s + t$, for subject i, denoted by $\pi_i(s, t)$, is defined as an estimate of:

$$P(T_i \leq s + t \,|\, T_i > s, \mathcal{H}_i(s)) \tag{1}$$

The longitudinal covariates $Y_i = \{Y_{ik}(t_{kij}); k = 1, ..., K; i = 1, ..., N, j = 1, ..., n_{ki}\}$ are typically **internal covariates**, which means that they are generated by the individual under study (e.g., blood pressure) by opposition with external covariates (e.g., level of air pollution)[5]. They can be built on longitudinal (bio)-markers (e.g., serum creatinine) or on the observation of related (intermediate) events (e.g., occurrence of disease complications when the aim is to predict death from the disease).

3 How are dynamic predictions computed?

Dynamic predictions are usually computed from a statistical model $P(\cdot; \theta)$ which depends on parameters θ. We here consider the case where an estimate of θ, denoted by $\hat{\theta}$, and an estimate

of the associated variance, denoted by $\widehat{V}(\hat{\theta})$, have already been obtained in a sample, and the aim is to compute dynamic predictions for new subjects.

A first possibility is to compute the dynamic predictions by plug-in. It means that for any new subject i, based on information $\mathcal{H}_i(s)$ the dynamic prediction can be computed as $\pi_i(s,t) = P(T_i \leq s+t|T_i > s, \mathcal{H}_i(s), \hat{\theta})$.

An alternative consists in approximating the distribution of the conditional probability by using the asymptotic distribution of the random parameter estimates. Indeed, $\hat{\theta}$ can be seen as the realization of a random variable which depends on the data on which the model has been estimated. This suggests a Bayesian approach and a Monte-Carlo computation: the principle is to draw a large number of θ_b ($b = 1, ..., B$) from the distribution $\mathcal{N}(\hat{\theta}, \widehat{V}(\hat{\theta}))$ and to compute dynamic predictions as $\pi_i(s,t) = \text{median}_{b\in 1,...,B} P(T_i \leq s+t|T_i > s, \mathcal{H}_i(s), \theta_b)$. This approach is especially useful to provide confidence bands for predictions such as those obtained from the 2.5% and the 97.5% percentiles of $P(T_i \leq s+t|T_i > s, \mathcal{H}_i(s), \theta_b)$ ($b = 1, ..., B$).[10].

The statistical models from which the predictions are computed are often called **prediction models**. They enable the mapping of the available subject-specific information onto a personalized risk of event. Different types of models can be used, especially depending on the kind of available information $\mathcal{H}_i(s)$ to use. Note that the estimation of prediction models most often needs to deal with censored observations that are frequently encountered with time-to-event data.

When the longitudinal information $\mathcal{H}_i(s)$ involves only time independent covariates, many standard time-to-event models can be used. Common models include Cox models and cause-specific (Cox) models in the presence of competing risks, but also others such as direct binomial regression models.

When the longitudinal information $\mathcal{H}_i(s)$ involves both time-independent covariates X_i and internal time-dependent covariates generated from a longitudinal (possibly multivariate) process Y_i, the statistical model needs to handle the dependency between Y_i and T_i. There are mainly two popular approaches to do so: the **joint modelling** and **landmarking** approaches[14].

Y_i and T_i can be modeled jointly in the so called **joint models** (cross-reference to stat06032). These models usually rely on a fully parametric approach which provides a thorough description of their dependency. However such a description is usually obtained at the price of potentially restrictive assumptions and can also lead to computationally demanding estimation procedures. Examples of dynamic predictions computed from joint models are provided by Proust-Lima and Taylor[10], Rizopoulos[11] when Y_i is a longitudinal marker and Mauguen et al.[6] when Y_i consists of recurrent or intermediate events. Note that when Y_i consists of intermediate events, a multistate approach for Y_i and T_i can alternatively be favored[4].

As only the conditional distribution of T_i given $T_i > s$ and $\mathcal{H}_i(s)$ is required for the computation of dynamic predictions, another option is, roughly, to fit standard time-to-event models on the subsample of patients still at risk at the **landmark** time. Baseline covariates are defined from the dynamic observations collected prior to the landmark time. See van Houwelingen and Putter[13], van Houwelingen[14] for a thorough description of this approach, often called **landmarking**, and Parast et al.[8], Nicolaie et al.[7] for derived dynamic predictions. On the one hand, this approach has the advantage that it avoids the need for specific assumptions induced by the full specification of the joint distribution of (Y_i, T_i). On the other hand, it requires a

careful summarization of the available dynamic information $\mathcal{H}_i(s)$ to create new baseline covariates to be included in the landmark models at each landmark time s. Note that within the conditional approaches other techniques also exist such as those based on partly conditional models[16].

4 How is dynamic prediction accuracy evaluated?

Whereas computing individual dynamic predictions is rather straightforward once a statistical model has been estimated, their practical utility remains unclear until their predictive accuracy has been properly evaluated.

Two key concepts of the predictive performance assessment are **calibration** and **discrimination**. Briefly, we say that a prediction model is **calibrated** if, for any percentage x, we can expect that x subjects out of 100 experience the event among all subjects that have a predicted probability of x %. Calibration is thus related to the concept of bias. Most of the recommended popular approaches to assess calibration simply consist in displaying the comparisons between observed and predicted risks within subgroups[9]. The **discrimination** concept is wider and not uniquely defined. It is however closely related to the variation of the predictions. The overall idea is that a prediction model has high discriminative power if (i) the range of the predictions it provides is wide and (ii) subjects with high (respectively low) predicted risk are more (respectively less) likely to experience the event. Among methods to assess discrimination, the most interesting and popular approaches are those derived from ROC curve analyses. They aim to compare the distribution of the predictions among subjects who experience the event of interest to the distribution of the predictions of subjects who do not[9].

Discrimination and calibration can also be evaluated at the same time using the quadratic prediction error, often called the **Brier score**[12]. Its scaled version provides an R^2-type criterion which measures how well predictions perform as compared to a reference null model, that is a model which does not include any subject-specific information. Although less attractive from the clinical interpretation point of view, it provides an interesting summary measure for a first glance at the prediction performances, especially for evaluating many models in order to select a small subset for further in-depth evaluation.

In the dynamic prediction context, prediction performances can be evaluated through the computation of time-dependent ROC curves[1], calibration plots and prediction errors for each landmark time s and a couple of time horizons t. One major challenge is however to summarize calibration and discrimination performances over the different landmark times and/or different time horizons: the more summarized the information the easier to communicate but the less informative. Good compromises are curves and tables displaying the evolution of prediction accuracy measures over different landmark times s for a fixed time horizon t[13]. For instance, curves displaying the area under the time-dependent ROC curve or the scaled Brier score evolution are sometimes considered[8,2].

Other more specific methods have also been proposed, for example the prognostic cross-entropy derived from the information theory for joint model selection[3].

5 Additional remarks

A few additional messages should be kept in mind when one is interested in dynamic predictions:

- Most of the time, the cause of an event is not unique. Dynamic predictions may need to be developed in a competing risks setting[2].
- A good model for etiology is rarely a good model for prediction and the other way around. The reason is that for models which aim to provide insights about etiology (e.g. causal models) the control of the bias should often be favored over the control of the variance. In contrast, controlling the variance is most important for prediction since "the clinical validation asks for robust approaches"[14].
- Estimation of measures of predictive accuracy described in Section 4 is not always straightforward in a dynamic prediction context. This is due to censoring and selection at the landmark time. One consensual approach for dealing with these issues is the inverse censoring probability weighting technique[12,8,2].
- Measures of predictive performances which are used to evaluate dynamic predictions often depend on the design of the study, which may complicate their clinical interpretation[15].
- Validation of predictive tools on external data is favored to avoid the over-optimism induced by the selection of the model on the same data, and the particularities of one dataset.
- Dynamic prediction is a new topic for which methodology is developing rapidly. Some techniques are however already implemented to assess dynamic predictions (e.g., R packages `pec`, `timeROC`) and/or compute dynamic predictions (e.g., R packages `dynpred`, `JM`, `JMBayes`, `lcmm`, `frailtypack` or Stata program `STJM`).

References

[1] Blanche, P., Dartigues, J.-F., and Jacqmin-Gadda, H. (2013). Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in medicine*, 32(30):5381–5397.

[2] Blanche, P., Proust-Lima, C., Loubère, L., Berr, C., Dartigues, J.-F., and Jacqmin-Gadda, H. (2015). Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics*, 71(1):102–113.

[3] Commenges, D., Liquet, B., and Proust-Lima, C. (2012). Choice of prognostic estimators in joint models by estimating differences of expected conditional Kullback-Leibler risks. *Biometrics*, 68(2):380–387.

[4] Cortese, G. and Andersen, P. K. (2010). Competing risks and time-dependent covariates. *Biometrical Journal. Biometrische Zeitschrift*, 52(1):138–158.

[5] Fisher, L. D. and Lin, D. Y. (1999). Time-dependent covariates in the Cox proportional-hazards regression model. *Annual Review of Public Health*, 20:145–157.

[6] Mauguen, A., Rachet, B., Mathoulin-Plissier, S., MacGrogan, G., Laurent, A., and Rondeau, V. (2013). Dynamic prediction of risk of death using history of cancer recurrences in joint frailty models. *Statistics in Medicine*, 32(30):5366–5380.

[7] Nicolaie, M. A., van Houwelingen, J. C., de Witte, T. M., and Putter, H. (2013). Dynamic Pseudo-Observations: A Robust Approach to Dynamic Prediction in Competing Risks. *Biometrics*, 69(4):1043–1052.

[8] Parast, L., Cheng, S.-C., and Cai, T. (2012). Landmark Prediction of Long Term Survival Incorporating Short Term Event Time Information. *Journal of the American Statistical Association*, 107(500):1492–1501.

[9] Pepe, M. and Janes, H. (2013). Methods for evaluating prediction performance of biomarkers and tests. In Lee, M.-L., Gail, G., Cai, T., Pfeiffer, R., and Gandy, A., editors, *Risk Assessment and Evaluation of Predictions*. Springer.

[10] Proust-Lima, C. and Taylor, J. M. G. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics (Oxford, England)*, 10(3):535–549.

[11] Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819–29.

[12] Schoop, R., Graf, E., and Schumacher, M. (2008). Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. *Biometrics*, 64:603–10.

[13] van Houwelingen, H. and Putter, H. (2011). *Dynamic Prediction in Clinical Survival Analysis*. Monographs on Statistics & Applied Probab #123. Chapman & Hall/CRC, London.

[14] van Houwelingen, H. C. (2014). From model building to validation and back: a plea for robustness. *Statistics in Medicine*, 33(30):5223–5238.

[15] Vickers, A. J. and Cronin, A. M. (2010). Everything you always wanted to know about evaluating prediction models (but were too afraid to ask). *Urology*, 76(6):1298–1301.

[16] Zheng, Y. and Heagerty, P. J. (2005). Partly Conditional Survival Models for Longitudinal Data. *Biometrics*, 61(2):379–391.