# Review of Gaussian copula approach for dynamic prediction of survival with a longitudinal biomarker

Kenny Chen *
Department of Mathematics and Statistics, Amherst College

October 22, 2021

**Abstract**

This paper summarizes the main points and methods presented in "A Gaussian copula approach for dynamic prediction of survival with a longitudinal biomarker" by Suresh et al. (2019) in the field of dynamic predictions in survival analysis. Dynamic prediction is the incorporation of information during patient follow ups to obtain updated predictions of their risk. The incorporation of updated biomarkers allows clinicians to respond appropriately to the patient's needs. Joint modeling and landmarking are two common approaches used for dynamic predictions; however, they have some disadvantages that might make them less attractive. Suresh et al. proposes an alternative method based on Gaussian copulas that aims to have the advantages of both approaches while avoiding their disadvantages. Their simulation studies show Gaussian copulas performed generally well when compared to joint modeling and landmarking and their application with real data further substantiates the performance of Gaussian copulas in dynamic prediction.

*Keywords:* Joint modeling, Landmarking, Survival analysis.

*The author gratefully acknowledges Amherst College, Nicholas Horton, Peers

# 1 Review

For my article review, I chose to review "A Gaussian copula approach for dynamic prediction of survival with a longitudinal biomarker" by Suresh et al. (2019) which proposed a copula based method for dynamic prediction that aimed to have the advantages of joint modeling and landmarking (two methods commonly used in dynamic survival prediction) while surpassing some of their limitations. They compared the predictive performance of their method versus two previously mentioned through simulation studies and a real world dataset.

In the introduction, Suresh et al. pointed out the increasing interest in survival analysis of patients beyond their baseline biomarker information. Estimation of risk for survival outcomes is usually obtained from a prediction model that uses information at the baseline. However, when we follow up with individuals, some biomarkers may be updated and this information ought to be incorporated into the survival prediction, thus producing what Suresh et al. calls "dynamic predictions". This is an important concept because dynamic predictions with updated biomarker information can then help clinicians make responsive and timely therapies or interventions.

To do this, it requires a technique that produces survival predictions at baseline and also incorporates additional biomarker information to produce updated predictions for patients still alive in the future. So dynamic predictions are seen as a conditional distribution ($[T|T > \tau, \boldsymbol{Z}(\tau)]$) where we have a model for failure time T that incorporates time-dependent biomarker information Z(t) where $Z(\tau)$ is the marker information available up to time $\tau$. The two most common methods to achieve this are "joint modeling" and "landmarking".

Joint modeling consists of two models: a model for Z(t) and a model for the failure time T. From these two models, a joint distribution [T,Z] can be derived. An advantage to joint modeling is that it produces a valid prediction function meaning we can get consistent conditional survival predictions. However, a disadvantage to joint modeling is that it requires full specification of the marker process which means making certain distributional assumptions about future values of the markers. Suresh et al. continued to say that the marker process may be hard to estimate with sparse longitudinal measurements and misspecifi-

cation of the model can lead to biased predictions. Moreover, it can be computationally burdensome for estimation and calculation of dynamic predictions.

On the other hand, landmarking requires specifying a survival model for $[T|T > \tau, \boldsymbol{Z}(\tau)]$ by using the empirical failure time distribution at certain fixed time points, $\tau$, conditional on being alive at $\tau$ and having marker information up until time $\tau$ (i.e. the marker value $Z(\tau)$). We estimate this empirical distribution using a Cox regression to model the hazard $h(t|\tau, Z(\tau))$. An advantage of landmarking is that it does not require the specification of a model for the marker process like in joint modeling; however, there are numerous decisions required in this method such as specifying beforehand the prediction times of interest (i.e. landmark times).

Suresh et al. then proposed a copula based approach for dynamic prediction that aimed to combine the advantages of joint modeling and landmarking while still maintaining good predictive performance. Copulas are multivariate cumulative distribution functions with uniform margins and through Sklar's theorem, we can link marginal distribution functions with the copula to create a joint distribution. This method requires specifying the marginal distribution functions $T|T > \tau$ and $Z|T > \tau$ for individuals alive at time $\tau$. Then a bivariate Gaussian copula models the joint distribution $((Z, T)|T > \tau)$. This approach allows us to specify the marginal distribution functions and model their association separately using the Gaussian copula.

In specifying models for the copula components, Suresh et al. prioritized simple flexible (possibly misspecified) models that approximate the true distribution. This avoided making strong modeling assumptions and allowed for easy estimation in software like R. They then go over how they chose to model the marker data, the failure time data, and the association. They further discussed that they chose the Gaussian copula for its tractable nature but recognized other copulas have different strengths and weaknesses.

Next, they demonstrated the performance of their method compared to joint modeling and landmarking through simulation studies. They measured performance using versions of area under the curve (AUC) and $R^2$ measures. Without going into the mathematical nuances, the idea is that they have true measures of AUC and $R^2$ and estimated versions so then smaller differences between the two indicate better performance. Furthermore,

they looked at the root mean squared prediction error (RMSPEs) between true conditional survival probabilities and predicted conditional survival probabilities to see if their method was consistently predicting higher or lower than the true probabilities.

They performed five hundred simulations, with 1000 subjects in each simulation, for various scenarios. Five hundred subjects were randomly sampled to create the training set while the rest made up the validation data set. They simulated patients who have been followed for 10 years with longitudinal biomarkers measured at baseline. They considered various patterns of biomarker observations over time as well as various models to assess robustness in more general situations. For joint modeling and landmarking models, they also considered various versions of these models which consisted of three joint modeling models and four landmark models. In addition, diagnostic plots and goodness-of-fit tests were used to examine the models.

After performing the simulation studies comparing all the various models, they concluded that the copula method had good predictive performance across the measures considered. Some key results they highlighted were that the copula method consistently performed better than the landmark model with baseline hazard as a function of landmark time and in certain situations, performed similarly to a joint model. Lastly, they found the copula method to be robust to the choice of the association function when the marginal models are well chosen. Results of one scenario are displayed below.

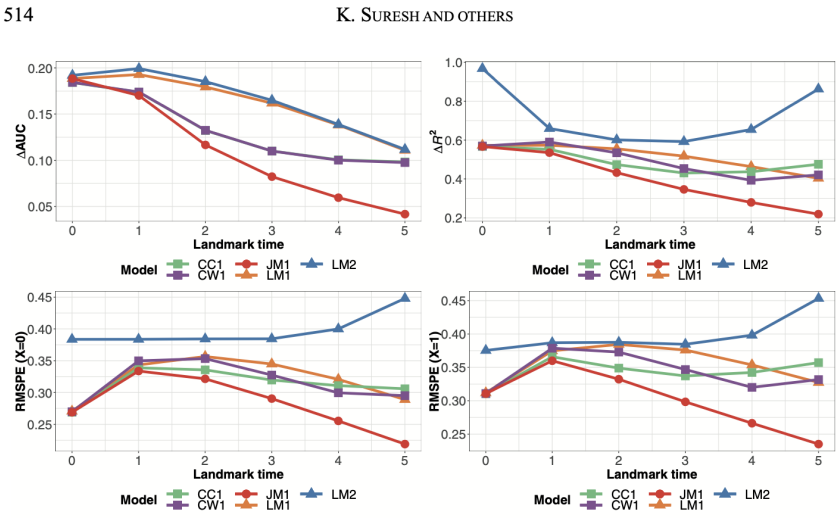


Fig. 1. Simulation estimates for Scenario 1a ($\sigma_\epsilon = 0.6$, $\phi = 1.5$, inspection rate 0.5) for $\Delta$AUC (top-left) and $\Delta R^2$ (top-right), and RMSPE for $X = 0$ (bottom-left) and $X = 1$ (bottom-right) for predicted probability $P(T \leq \tau + 3 | T > \tau, Z(\tau), X)$ from copula models (CC1), (CW2), joint model (JM1), and landmark models (LM1), (LM2). Lower values indicate better predictive performance.

Next, they compared the methods using real world data in the form of an aortic heart valve study. They used data from an observational study that consisted of 248 patients who had gotten an aortic valve replacement to compare the efficacy of two types: homograft or stentless. Longitudinal measurements of the left ventricular mass index (LVMI) were measured at baseline (after surgery). Build up of mass in the left ventricular can lead to heart attacks which incentivizes measuring individual's changing LVMI to perform dynamic predictions of risk of death. They utilized 5-fold cross-validation to compute the measures used in the simulation study and found that all the methods performed similarly.

Lastly, they discussed some of the limitations to their proposed method. They noted that it relies on the availability of data at prediction times of interest to properly model the joint distribution. Because the marginals are models and there is a model for the association, a lot of parameters need to be estimated. Moreover, Gaussian copulas only apply when linking two continuous outcomes.

In summary, this paper proposed a new method for dynamic predictions as an alternative to two common methods, joint modeling and landmarking. They proposed a copula-based approach that aimed to have the advantages of both joint modeling and landmarking while mitigating the cons of both. They then performed simulation studies to compare the three methods and found that their proposed method generally had good predictive performance. And when they applied all three methods to an aortic valve study, they found all methods gave similar results.

Overall, as a student in STAT495, I think this paper can appeal to a broad audience. For those who might not know much about this subject (like myself), I felt it was comprehensible and I could grasp the big ideas while at the same time, it also had more theoretical information for those more educated in this subject. The information was clear and it was well organized.

# References

Suresh, K., Taylor, J. M. G. & Tsodikov, A. (2019), 'A Gaussian copula approach for dynamic prediction of survival with a longitudinal biomarker', *Biostatistics* **22**(3), 504–

521.

**URL:** *https://doi.org/10.1093/biostatistics/kxz049*