Dynamic Prediction of Disease Progression Using Longitudinal Biomarker Data

Xuelin Huang

Department of Biostatistics

M. D. Anderson Cancer Center

The University of Texas

Joint Work with

Jing Ning, Sangbum Choi, Alfonso Quintas-Cardama, and Jorge Cortes

Outline

1. What is Dynamic Prediction?
2. A Motivating Example: Chronic Myelogenous Leukemia (CML)
3. Current Methods
   - Joint Modelling of Longitudinal and Survival Data: Not well-suited for prediction
   - Landmark Analysis: Separate unrelated predictions on discrete time points
4. Proposal: An Information-cumulating Model for Predictive Analysis Continuously over Time (IMPACT)
5. Dynamic Predictive Analysis for CML

Dynamic Prediction

- Keep making updated predictions as time goes by and more data are observed
- After treatment, we need prediction of future disease prognosis at all the time points during a patient's follow-up visits.
- To decide whether or not to initiate extra treatments or interventions.
- Need use not only the baseline information, but also all the information up to the time point of prediction.

Chronic Myelogenous Leukemia (CML)

- The first human cancer that was linked to a single, acquired abnormal gene, the BCR-ABL gene.
- Tyrosine kinase inhibitors (TKIs) can inhibit the BCR-ABL gene.
- Frontline treatment trial of TKIs was usually successful: motivating data set for this talk.
- TKIs are not chemotherapy, have no severe side effects.
- The disease residual can be measured by the expression level of the BCR-ABL gene.

Chronic Myelogenous Leukemia (CML)

- Patients have their BCR-ABL expression levels measured roughly every three months, but in reality can be any time.
- Current practice is to wait until disease relapse (with clinical symptoms) to initiate other treatments
- Question: Can we use BCR-ABL levels to predict future disease relapse and initiate other treatments for early prevention?

Chronic Myelogenous Leukemia (CML)

- **Note: An increasing of BCR-ABL during prolonged remissions does not automatically constitute relapse on its own. Reasons:**
  - **Patient's failure to comply (the pills are expensive, need to take everyday)**
  - **BCR-ABL trajectories have cyclic oscillations**
- **Initiating other treatments too early is not good either, because they are toxic and risky chemotherapies / stem cell transplant**
- **Need a good dynamic prediction model**

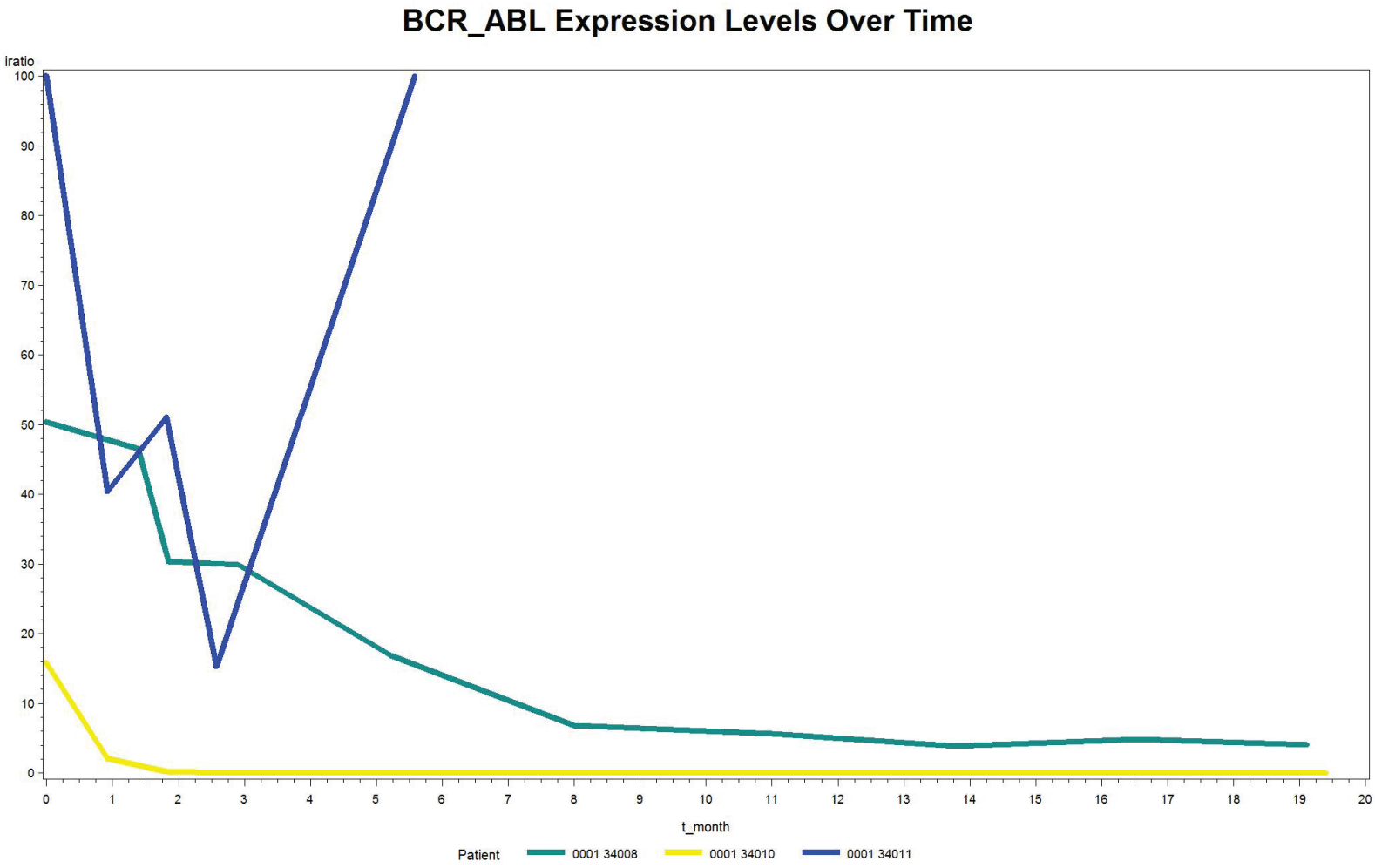


Figure 1: Biomarker Trajectories for Three Patients

Chronic Myelogenous Leukemia (CML)

- Biomarker changing patterns vary greatly from patient to patient
- It is difficult to use parametric models to fit such longitudinal data
- No, I am not going to use non-parametric models
- I will try to avoid using a longitudinal model for biomarker data
- Still, I need to use longitudinal biomarker data to predict survival

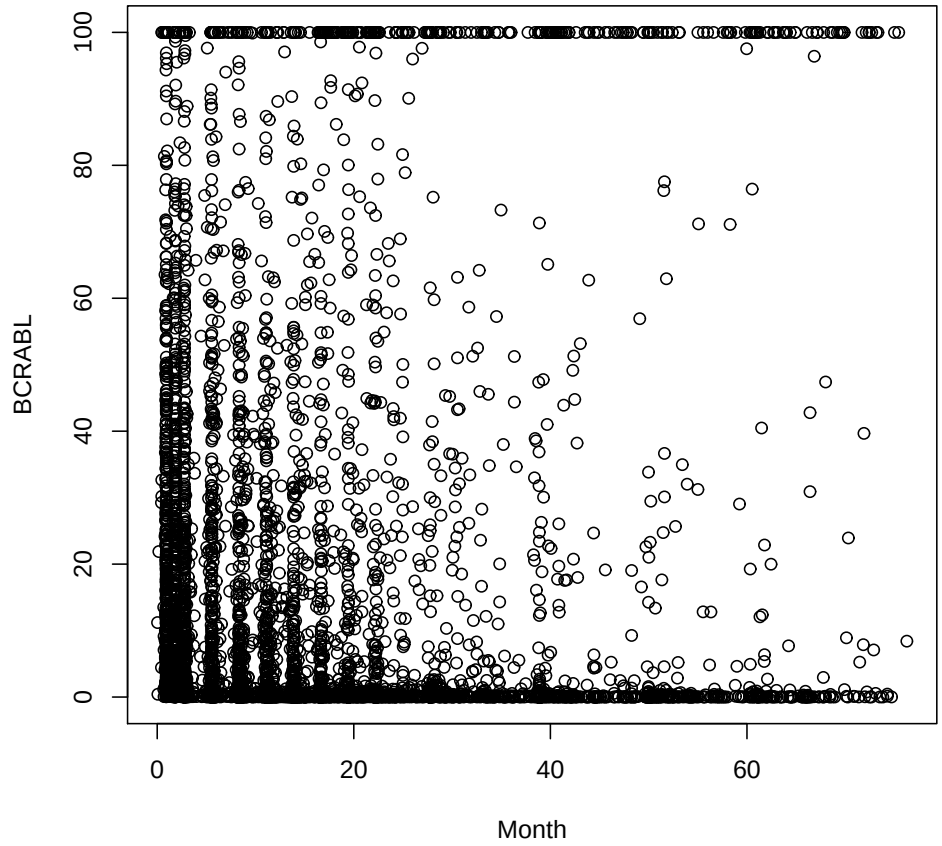


Figure 2: BCR-ABL Measurements for All Patients

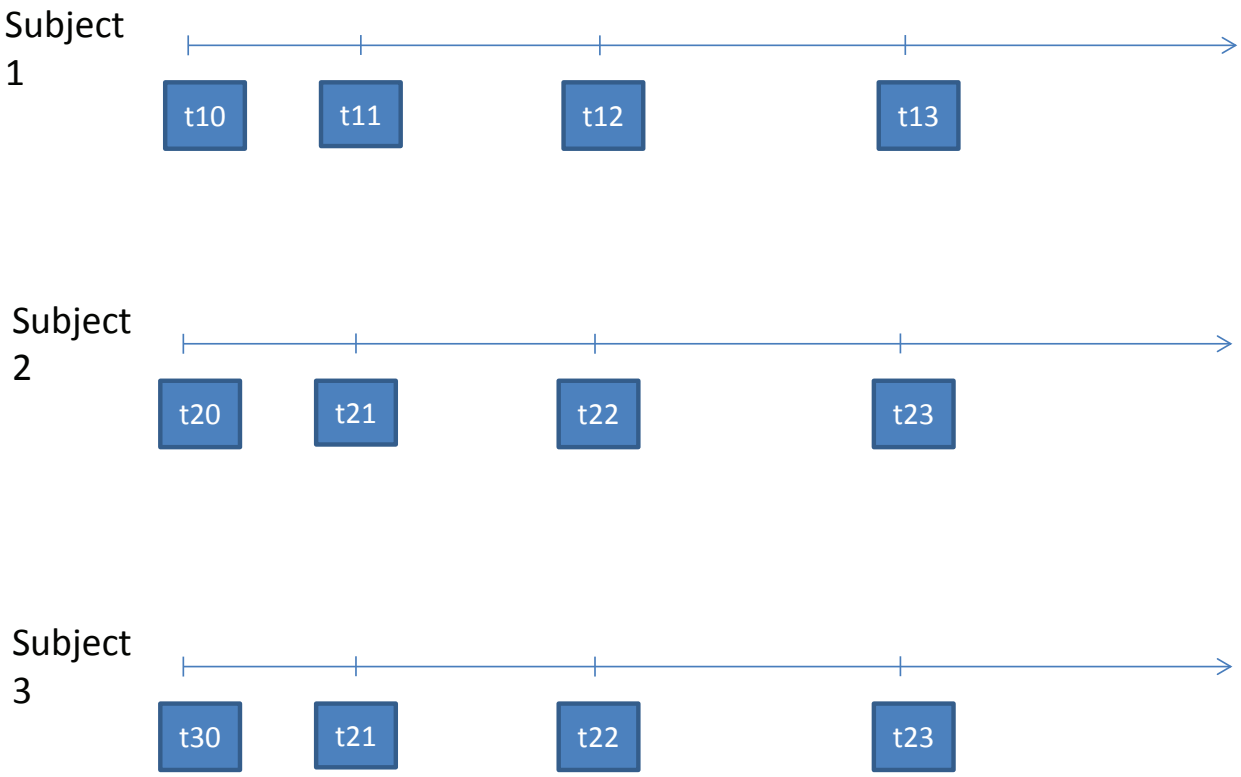


Figure 3: Regular repeated measurements for biomarkers

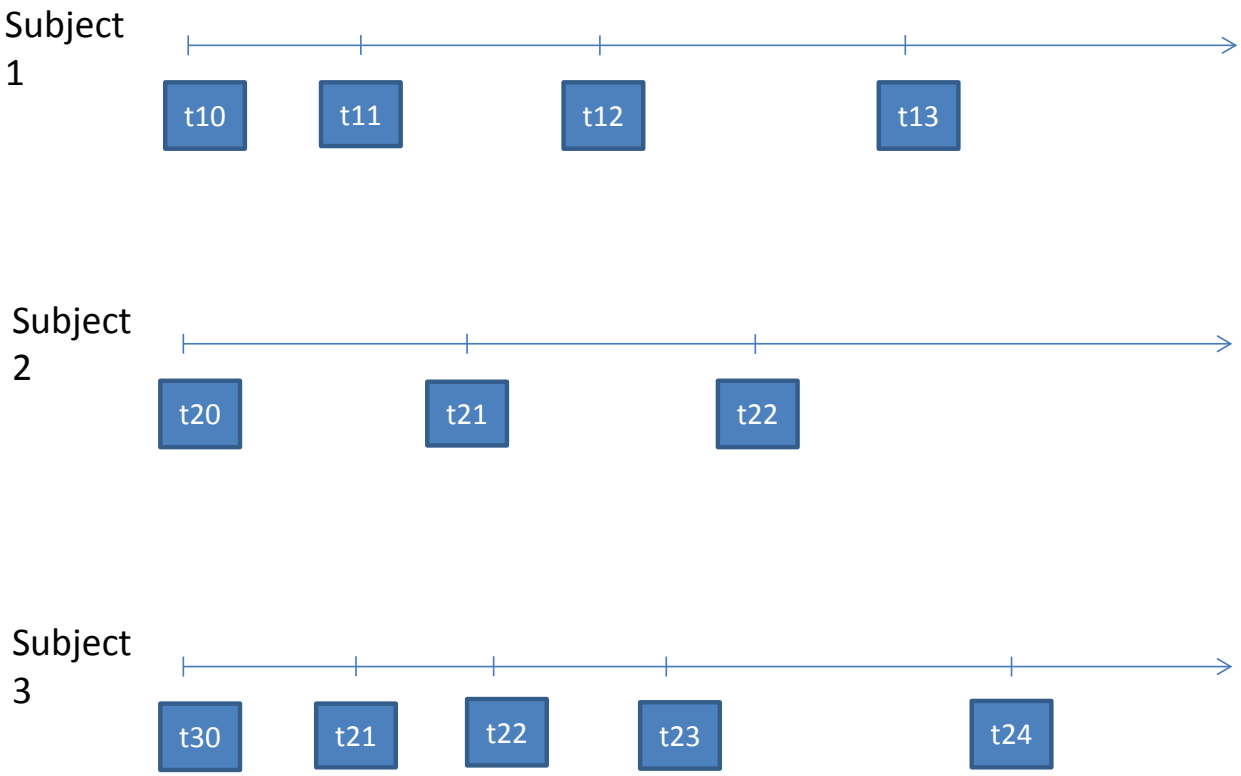


Figure 4: Irregular repeated measurements for biomarkers

Chronic Myelogenous Leukemia (CML)

- Need use BCR-ABL expression level to predict future disease relapse
- Patients may visit any time between the scheduled visits, so need do prediction at any time, not just some specific time points
- Prediction model should be able to use biomarker measurements from irregular time intervals

Notation

T_i: Time to disease relapse, or simply survival time

C_i: Censoring time

$X_i = \min(T_i, C_i)$, $\Delta_i = T_i \leq C_i$

$\lambda_i(t)$: Hazard function of T_i, describing failure risk rate at time t

Y_i: Baseline covariates

$Z_i(t)$: longitudinal biomarker value at time t

t_{ik}: the k^{th} biomarker measurement time for the i^{th} subject, $k = 1, \cdots, n_i$.

Current Approaches for Dynamic Prediction

1. Joint modeling of longitudinal biomarkers and survival data
2. Landmark analysis

Joint Modeling: Current Approach (1) for Dynamic Prediction

Joint modeling of longitudinal biomarkers and survival data

- **Use random effect model for longitudinal data**
- **Cox proportional hazards model for survival, with longitudinal biomarkers as time-dependent covariates**

Current Approach (1) Joint Modeling: Inconvenience for Prediction

Model: $\lambda_i(t) = \lambda_0(t) \exp\{\beta' Z_i(t)\}$,

Prediction at time t, conditional on $T_i \geq t$,

$$\begin{aligned} & \Pr(T_i \geq t+v | T_i \geq t) \\ = \quad & \exp\left[- \int_t^{t+v} \lambda_0(u) \exp\{\beta' Z_i(u)\} \, du \right] . \end{aligned}$$

Inconvenience (1):

Need future values of $Z(u)$ for $u > t$ that are not available yet at the time t.

Current Approach (1) Joint Modeling: Inconvenience for Prediction

Assume event times $\{x_i : i = 1, \cdots, n\}$ sorted ascendingly without ties, need maximize $L(\beta)$ to estimate β,

$$L(\beta) = \prod_{i=1}^{n} \left[\frac{\exp\{\beta' Z_i(x_i)\}}{\sum_{j \geq i} \exp\{\beta' Z_j(x_i)\}} \right]^{\Delta_i}$$

Inconvenience (2):

For each event time x_i, need not only $Z_i(x_i)$, but also $Z_j(x_i)$ for all $j \geq i$.

Such $Z_j(x_i)$ are usually not observed.

Current Approach (2)
Landmark Analysis for Dynamic Prediction

- Do predictions at only some selected time points.
- For each selected time point, use a Cox model with only time-independent covariate to summarize biomarker information up to this point.
- Does not use information after this point, i.e., no need to use future biomarker values.

Current Approach (2)
Landmark Analysis: Inconveniences

- **Can be done only at selected time points**

$$\begin{aligned}
\lambda_{i,0}(t) &= \lambda_{0,0}(t)\exp\{\beta_0' Z_i(0)\}\,, \\
\lambda_{i,3}(t) &= \lambda_{0,3}(t)\exp\{\beta_3' Z_i(3)\}\,, \\
\lambda_{i,6}(t) &= \lambda_{0,6}(t)\exp\{\beta_6' Z_i(6)\}\,, \\
&\cdots\cdots
\end{aligned}$$

- **Over-parameterized with $\lambda_{0,0}(t), \lambda_{0,3}(t), \lambda_{0,6}(t), \cdots$ and $\beta_0, \beta_3, \beta_6, \cdots$.**
- **Smoothing techniques have been used to put constraints on the above parameters**

References

- Tsiatis and Davidian (2001): A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error.
- Zheng and Heagerty (2005): Partly conditional survival models.
- van Houwelingen (2007), van Houwelingen and Putter (2008): Dynamic prediction by landmarking in event history analysis.
- Putter et al (2007): Competing risks and multi-state modeling

A New Approach for Dynamic Prediction

We try to provide a method that

- **does prediction at any time point, not just on pre-specified time points such as $t = 0, 3, 6, \cdots, t_m$.**
- **does not use future value $Z(t + v)$ for prediction at time t**
- **does not need a model for covariates**
- **can use biomarker measurements from irregular time intervals**
- **dose not need to fill biomarker values on other subjects' event time points.**

Proposal: Information-cumulating Model for Predictive Analysis Continuously over Time (IMPACT)

The new approach

- **is modified from landmark analysis**
- **so does not need to use future value $Z(t+v)$ for prediction at time t**
- **Landmark analysis fits m separate models, one for each selected time point.**
- **The new approach uses two-stage modeling,**
  - **1st stage: Fit a Cox model for $t = 0$,**
  - **2nd stage: Add on to the model for $t = 0$ to fit for all $t > 0$.**

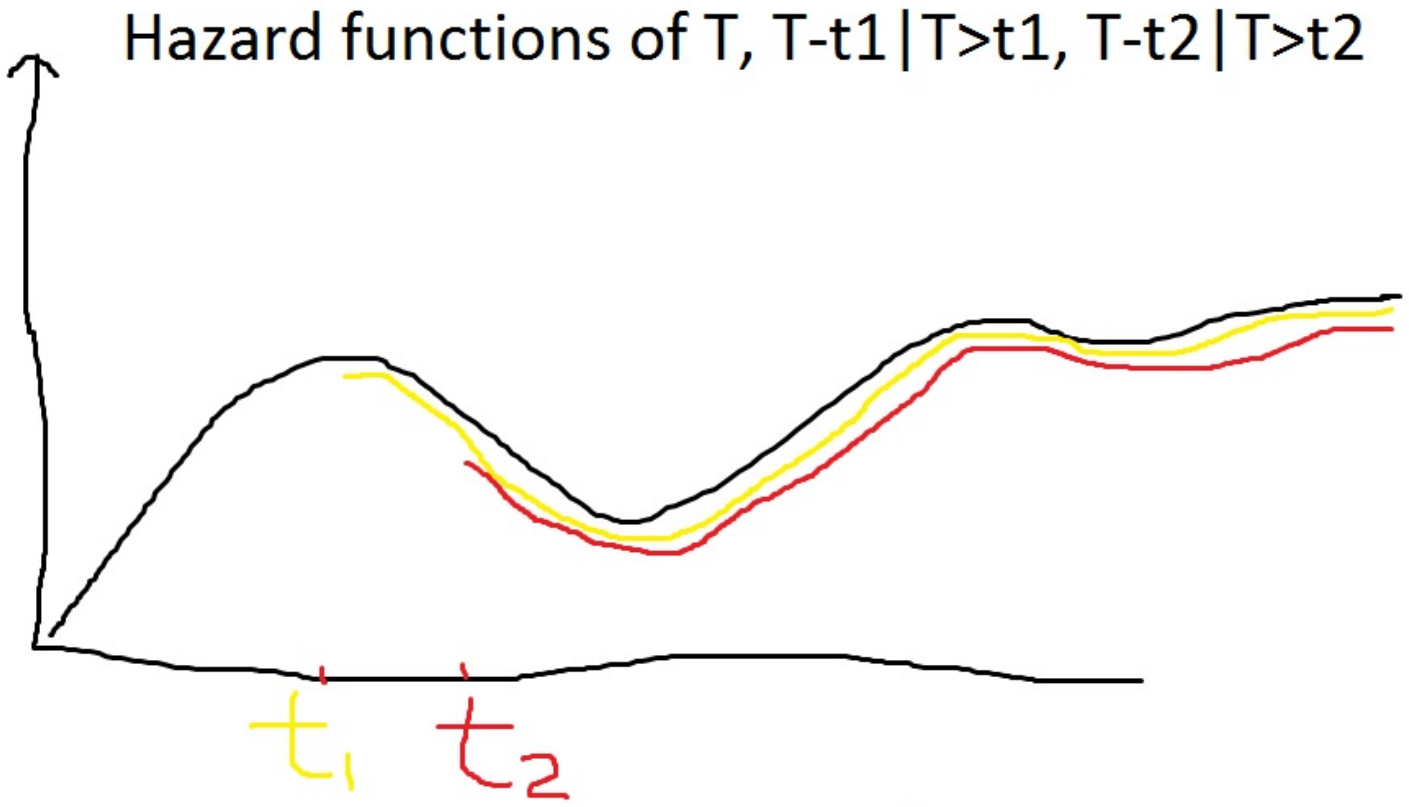
Information-cumulating Model for Predictive Analysis Continuously over Time (IMPACT)

Key step: How to add on to the model for $t = 0$ to fit for all $t > 0$?

Answer: Use a fundamental equality for conditional survival.

Suppose $\lambda_0(u)$ is the hazard function for T,
and $\lambda_t(u)$ is the hazard function for $T - t | T > t$ for $T - t = u$.

Then we have $\lambda_t(u) = \lambda_0(t + u)$ for all $t > 0$.


Hazard functions of T, T-t1|T>t1, T-t2|T>t2
t1
t2

Derivation (1)

Let $S_0(t) = \Pr(T \geq t) = \exp(-\int_0^t \lambda_0(v)\, dv)$, then,

$$\Pr(T \geq t+u | T \geq t) = \frac{S_0(t+u)}{S_0(t)}$$

$$\triangleq \quad S_t(u) = \exp(-\int_0^u \lambda_t(v)\, dv)$$

$$= \quad \frac{\exp(-\int_0^{t+u} \lambda_0(v)\, dv)}{\exp(-\int_0^t \lambda_0(v)\, dv)} = \exp(-\int_t^{t+u} \lambda_0(v)\, dv)$$

$$= \quad \exp(-\int_0^u \lambda_0(t+v)\, dv)$$

$\Longrightarrow \lambda_t(v) = \lambda_0(t+v)$, i.e., $\lambda_3(v) = \lambda_0(v+3), \cdots$.

Derivation (2)

$$
\begin{aligned}
\lambda_0(v) &= \frac{-S'(v)}{S(v)}\,, \\
\lambda_t(v) &= \frac{-S_t'(v)}{S_t(v)} \\
&= \frac{\frac{-\partial S(t+v)/S(t)}{\partial v}}{S(t+v)/S(t)} \\
&= \frac{-\frac{\partial S'(t+v)}{\partial v}}{S(t+v)} \\
\Longrightarrow \lambda_t(v) &= \lambda_0(t+v)
\end{aligned}
$$

Information-cumulating Model for Predictive Analysis Continuously over Time (IMPACT)

- **Fundamental equality:** $\lambda_t(v) = \lambda_0(t+v)$.
- **Use this inherent constraint for hazard functions of the same survival time** T **at different time origins.**
- **Result in a more parsimonious approach for prediction of** T **at any** $t > 0$ **given** $T \geq t$.

Information-cumulating Model for Predictive Analysis Continuously over Time (IMPACT)

- **Stage 1: Using only demographics and biomarker information at baseline ($t = 0$) for prediction**
- **Stage 2: Using longitudinal biomarker information beyond baseline ($t > 0$) to improve prediction obtained from stage 1 (Information-cumulating)**

Information-cumulating Model for Predictive Analysis Continuously over Time (IMPACT)

Stage 1: Use a Cox model with only baseline (time-independent) covariates Y_i

$$\lambda_i(t) = \lambda_0(t) \exp\{\alpha' Y_i\} ,$$

This implies, without using any longitudinal data beyond baseline, prediction at time t can be done by

$$\begin{aligned} & S_i(t+u|T_i \geq t, Y_i) \\ = & \frac{S_i(t+u|Z_i)}{S_i(t|Y_i)} = \left\{ \frac{S_0(t+u)}{S_0(t)} \right\}^{\exp(\alpha' Y_i)} . \end{aligned}$$

Information-cumulating Model for Predictive Analysis Continuously over Time (IMPACT)

Stage 2: At time t, with longitudinal data $Z_i(t)$, postulate the hazard function of $T_i - t$ as

$$\begin{aligned} \lambda_{i,t}(u) &= \lambda_{0,t}(u)\exp\{\alpha' Y_i + \beta'(t) Z_i(t)\} \\ &= \lambda_0(t+u)\exp\{\alpha' Y_i + \beta'(t) Z_i(t)\}\,. \end{aligned}$$

Notes:

- **Infinite number of reference hazard functions $\lambda_{0,t}(u)$ indexed by $t > 0$ have been expressed by a single reference hazard function $\lambda_0(t+u)$.**
- **Need smoothness assumptions for $\beta(t)$**

Information-cumulating Model for Predictive Analysis Continuously over Time (IMPACT)

Then the previous prediction

$$\Pr(t+u|T_i \geq t, Y_i) = \left\{ \frac{S_0(t+u)}{S_0(t)} \right\}^{\exp(\alpha' Y_i)} .$$

can be improved by

$$\begin{aligned} & \Pr(T_i \geq t+u|T_i \geq t, Y_i, Z_i(t)) \\ = & \frac{S_i(t+u|Z_i(t))}{S_i(t|Y_i, Z_i(t))} = \left\{ \frac{S_0(t+u)}{S_0(t)} \right\}^{\exp\{\alpha' Y_i + \beta'(t) Z_i(t)\}} \end{aligned}$$

with improvement achieved by additional information in $Z(t)$.

Information-cumulating Model for Predictive Analysis Continuously over Time (IMPACT)

- Longitudinal data $Z_i(t)$ are used to further distinguish subjects surviving at time t.
- Subjects may have $\beta'(t)Z_i(t) > 0, \ = 0$, or < 0
- Correspond to prediction by using $Z(t)$ being worse, equal or better than prediction without using $Z(t)$.

Information-cumulating Model for Predictive Analysis Continuously over Time (IMPACT)

- Note Stage 2 specifies a landmark analysis model for each $t > 0$.
- Recall that landmark analysis does not use future values for prediction.
- This is why the new approach does not need use future biomarker data in prediction.
- Next a few slides show how we avoid using unobserved $Z_j(x_i), j \geq i$.

Two-Stage Parameter Estimation for IMPACT

Stage 1: Estimate α and $S_0(t), t \geq 0$.

- **Only the baseline covariate Y and survival information are used to fit a Cox model (with time-independent covariates).**
- **Maximizing partial likelihood to obtain $\hat{\alpha}$**
- **The Breslow estimator for $S_0(t), t \geq 0$.**

$$\hat{S}_0(t) = \exp\left\{ -\sum_{x_i \leq t} \frac{\delta_i}{\sum_{x_j \geq x_i} \exp(\hat{\alpha}' Y_j)} \right\}$$

Two-Stage Parameter Estimation for IMPACT

Stage 2: Estimate $\beta(t)$

- **From a subject with data Y, $Z(t_1)$, $Z(t_2)$, $\cdots$, $Z(t_m)$ and survival T,**
- **Create m pseudo-subjects with data shown below:**

  Subject 1: Baseline covariates Y and $Z(t_1)$, survival time $T - t_1$;

  $\cdots$

  Subject m: Baseline covariates Y and $Z(t_m)$, survival time $T - t_m$;

- **Each pseudo-subject contributes a likelihood term.**

Two-Stage Parameter Estimation for IMPACT

- **Each pseudo-subject contributes a likelihood term.**

  Subject 1**:** $T - t_1 \sim \left\{ \frac{\hat{S}_0(t_1+t)}{\hat{S}_0(t_1)} \right\}^{\exp\{\hat{\alpha}' Y_i + \beta'(t_1) Z(t_1)\}}$;

  $\ldots$

  Subject m**:** $T - t_m \sim \left\{ \frac{\hat{S}_0(t_m+t)}{\hat{S}_0(t_m)} \right\}^{\exp\{\hat{\alpha}' Y_i + \beta'(t_m) Z(t_m)\}}$.

Two-Stage Parameter Estimation for IMPACT

Stage 2: Estimate $\beta(t)$ (re-parameterized into β)

- **Working independence between pseudo-subjects**
- **Pseudo-likelihood = product of likelihood terms of all pseudo-subjects**
- **Maximize pseudo-likelihood to estimate β,**
- **With $\hat{\alpha}$ and $\hat{S}_0(\cdot)$ being fixed in Stage 2.**
- **Fixed $\hat{S}_0(\cdot)$ eliminates the need to use Cox-type partial likelihood for estimating β, and so eliminates the need to know $Z_j(x_i), j \geq i$.**
- **The only unknown parameter in the pseudo likelihood is β.**

Information-cumulating Model for Predictive Analysis Continuously over Time (IMPACT)

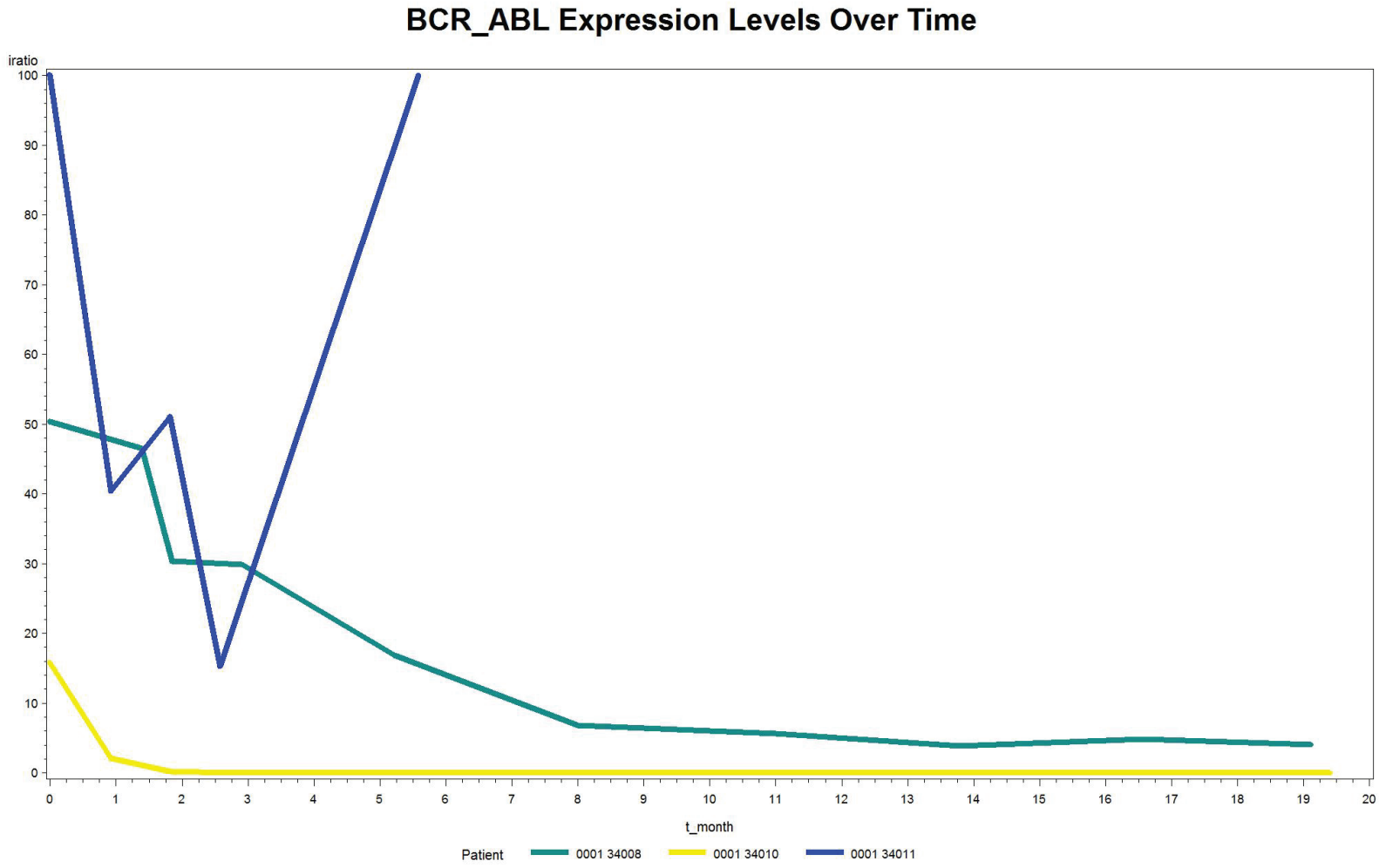
- **Using a training data set, get estimators $\hat{\alpha}$, $\hat{S}_0(t), t \geq 0$ (Stage 1), and $\hat{\beta}(t)$ (Stage 2).**
- **For a new subject, at time t with covariate value $Z_{\text{new}}(t)$, predict his survival distribution as**

$$\begin{aligned} & \Pr(T_{\text{new}} \geq t+u | T_{\text{new}} > t, Y_{\text{new}}, Z_{\text{new}}(t)) \\ = & \left\{ \frac{\hat{S}_0(t+u)}{\hat{S}_0(t)} \right\}^{\exp\{\hat{\alpha}' Y_{\text{new}} + \hat{\beta}'(t) Z_{\text{new}}(t)\}} \end{aligned}$$

Information-cumulating Model for Predictive Analysis Continuously over Time (IMPACT)

Assume a parametric form or use splines for $\beta(t)$.

- Trade-off between
  - Using parametric models for $Z(t)$ to impute covariate values at time points they are not observed
  - Assuming a parametric form for $\beta(t)$.
- It is reasonable to believe that the true shape of $\beta(t)$ is more smooth than covariate $Z(t)$.
- Covariate $Z(t)$'s are very bumpy, see next.


BCR_ABL Expression Levels Over Time
iratio
100
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t_month
Patient
0001 34008
0001 34010
0001 34011

Information-cumulating Model for Predictive Analysis Continuously over Time (IMPACT)

An example of a parametric form $\beta(t)$, after re-parameterizing,

$$
\begin{aligned}
& \Pr(T_i \geq t+u | T_i \geq t, Z_i(t)) \\
= & \frac{S_i(t+u|Z_i(t))}{S_i(t|Z_i(t))} \\
= & \left\{ \frac{S_0(t+u)}{S_0(t)} \right\}^{\exp\{\alpha' Y_i + \beta_0' Z_i(0) + \beta_1' \ln(t+1) Z_i(t)\}}
\end{aligned}
$$

CML Example

The model for dynamic prediction

$$\begin{aligned} & \Pr(T_{\textbf{new}} \geq t+u | T_{\textbf{new}} > t, Z_{\textbf{new}}(t)) \\ = & \left\{ \frac{\hat{S}_0(t+u)}{\hat{S}_0(t)} \right\}^{\exp\{\hat{\alpha}' Y_{\textbf{new}} + \hat{\beta}'(t) Z_{\textbf{new}}(t)\}} \end{aligned}$$

with

$$\begin{aligned} & \hat{\alpha}' Y_{\textbf{new}} + \hat{\beta}'(t) Z_{\textbf{new}}(t) \\ & = 0.458\, I(\textbf{age} > 60) + 0.0185\, \textbf{BCR}(0) \\ & \quad -0.298 \log(t+1) - 0.002\, \textbf{BCR}(t)\ \log(t+1). \end{aligned}$$

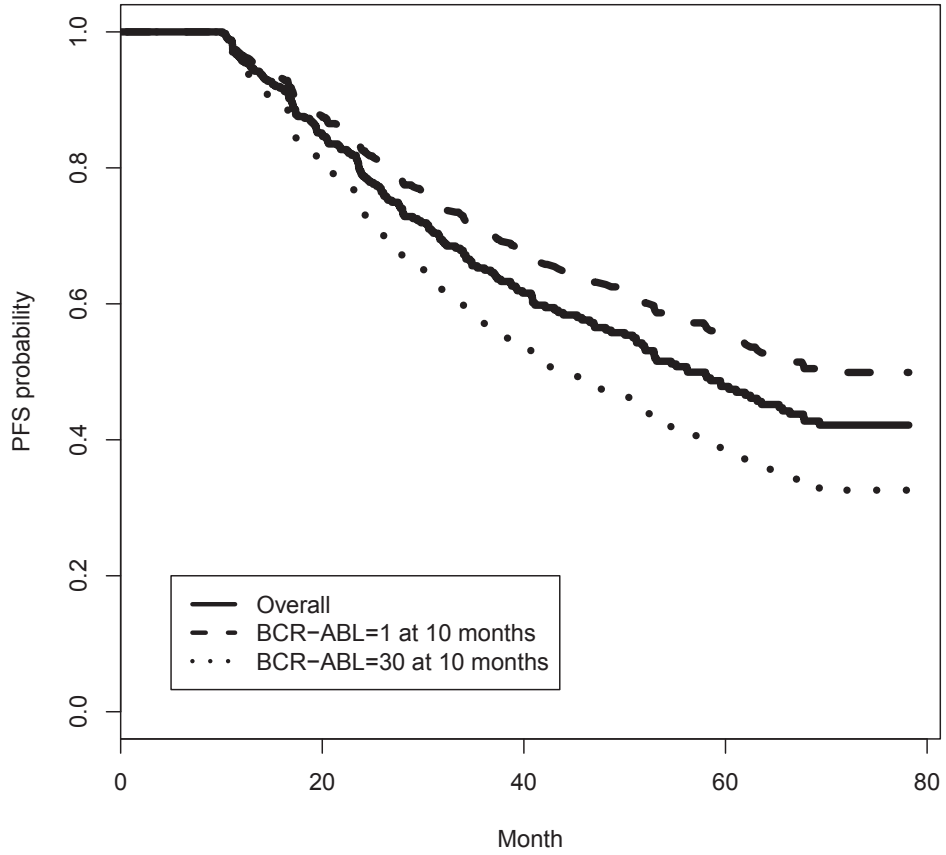


Figure 5: (1) Without using Z(10) (solid line), (2) Z(10)= 1 (dashed line), and (3) Z(10)= 30 (dotted line).

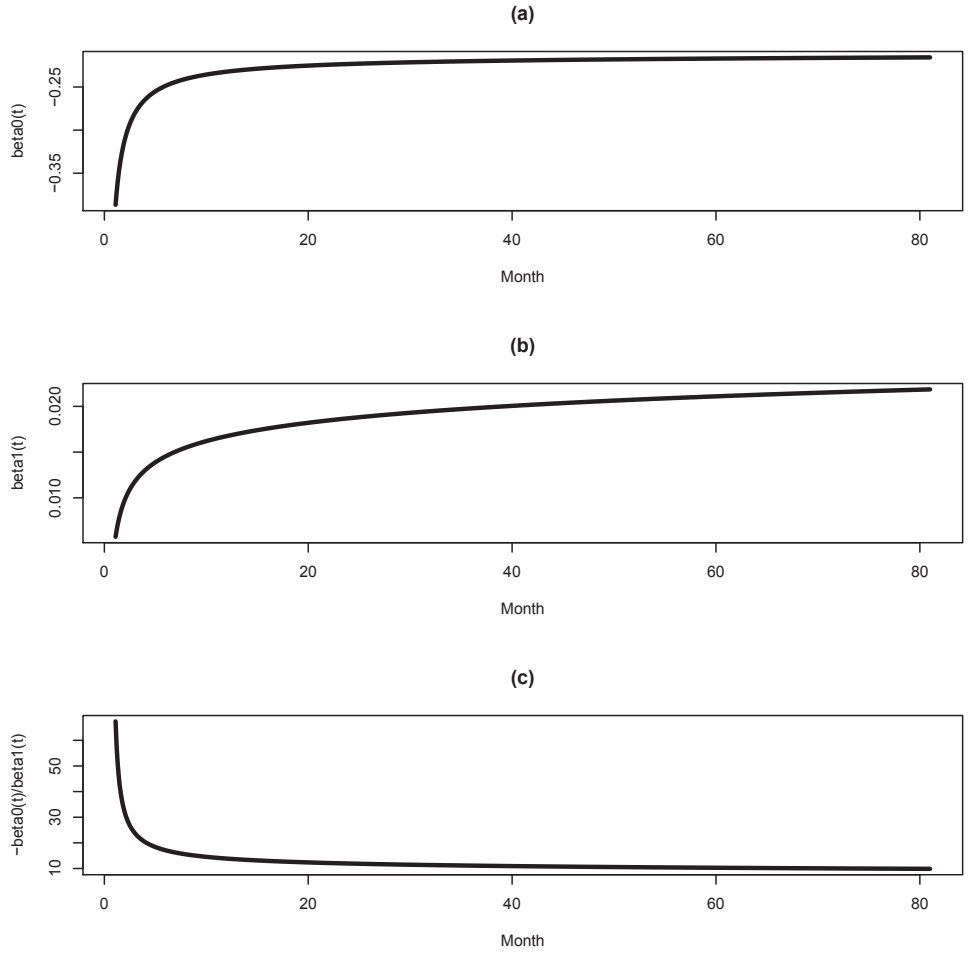


Figure 6: A biomarker trajectory with average (typical) survival

Summary

- Proposed approach uses a series of landmark analysis models over continuous t that are smoothed by using a parametric or spline $\beta(t)$.
- Landmark analysis approach avoids need to use future biomarker values in prediction.
- Two-stage estimation approach
  - avoids need of $Z_j(x_i), j \geq i$ by estimating $\lambda_0(t)$ (and thus $S_0(t)$) from the 1st stage and being fixed at 2nd stage
  - avoids need of model for $Z(t)$.

Discussion

- The estimation and interpretation of α are not distorted by intermediate outcomes reflected in time-dependent covariates $Z(t)$.
- This is usually what we want, i.e., α estimates the marginal population effects of baseline covariates on survival.
- The interpretation of corresponding regression coefficients in joint modeling is awkward.
- Bottom line: Proposed approach is easy and convenient to use.