Published in final edited form as:

Biometrics. 2017 September ; 73(3): 742–744. doi:10.1111/biom.12630.

Discussion of “A risk-based measure of time-varying prognostic discrimination for survival models,” by C. Jason Liang and Patrick J. Heagerty

Layla Parast* and Carolyn M. Rutter

RAND Corporation, 1776 Main Street, Santa Monica, CA 90401

Abstract

We congratulate Drs. Liang and Heagerty (L&H) for introducing a risk-based measure of time-varying discrimination for determining the value of predictors in time-to-event models. Their proposed hazard discrimination summary (HDS) is intuitive and its statistical properties are appealing. The local-in-time model framework provides additional flexibility for departures from the Cox proportional hazards model. Our discussion focuses on use of the proposed method to determine the added value of intermediate event information and an alternative approach for confidence interval estimation.

1 DETERMINING THE VALUE OF INTERMEDIATE EVENT (IE) INFORMATION

In clinical settings, information about an intermediate (or short term) event related to the primary (long term) outcome of interest can be used to update prediction after a pre-specified time to increase the predictive accuracy. We consider an application of L&H’s HDS that incorporates IE information, based on a landmark model framework (van Houwelingen & Putter, 2011; Parast et al., 2012). Methods for assessing the incremental value of IE information have focused on using the area under the receiver operating characteristic curve (AUC) and cumulative risk. The HDS offers an alternative approach to assess the incremental value of IE information. This new perspective offers the potential to provide insights into the effect of an IE on survival.

Let T_{Si} denote the time of the intermediate event for individual i . We assume semi-competing risks, where T_{Si} can be censored by both C_i and T_i . Let $Y_{Si} = \min(T_{Si}, T_i, C_i)$ and $\delta_{Si} = I(T_{Si} < \min(T_i, C_i))$. For example, in patients with end stage renal disease who are initiating dialysis, hospitalization within the first 6 months of dialysis initiation is an important predictor of 1-year survival (Parast et al., 2011; Parast et al., 2012). Because our interest is in predicting events after the landmark time, t_0 , we develop a prediction rule using the subset of individuals observed to survive past t_0 , $\Omega_{t_0} = \{Y_i > t_0\}$. Let

$M_i^* = \{M_i, I(T_{Si} \leq t_0), \min(T_{Si}, t_0)\}$ be the vector of predictors available at t_0 for individual i .

Within the subset $\Omega(t_0)$, we can measure discrimination by

* parast@rand.org.

$$HDS(t | \Omega(t_0)) = \frac{E_{M^*} \{ \lambda(t | M^*, t_0) | T = t, T > t_0 \}}{E_{M^*} \{ \lambda(t | M^*, t_0) | T > t, T > t_0 \}}.$$

$HDS(t | \Omega(t_0))$ can be estimated using the approach proposed by L&H applied to the subset of individuals in Ω_{t_0} , and using M_i^* instead of M_i in equations (2) and (3) of their paper. Note

that use of a model-based estimation approach such as the Cox model does require that the landmark model holds i.e. $\lambda(t | M, T > t_0) = \lambda_0(t | T > t_0) \exp(\beta^{*'} M^*)$ (or similarly, a local-in time version), in addition to the assumptions regarding the true survival model required by L&H. If the model assumptions are not likely to hold, alternative methods that are not within the Cox model framework could be considered (Parast et al., 2011; Parast et al., 2012).

The ratio $HDS(t | \Omega(t_0), M^*) / HDS(t | \Omega(t_0), M)$ captures the incremental value of the intermediate event information among those surviving to the landmark. We illustrate the potential utility of the HDS quantity by simulating an intermediate event for individuals in the Mayo PBC dataset. For example, such an event could be response to ursodeoxycholic acid (UDCA) within the first 12 months of treatment initiation which has been shown to be a strong predictor of prognosis in patients diagnosed with PBC (Carey et al., 2015). We simulate three scenarios based on the PBC data, one with a non-informative IE, one with a mildly informative IE and another with a moderately informative IE. To simulate uninformative IE times, we let $T_{Si} = \exp(N_i)$ where $N_i \sim \text{Normal}(5.5, 4)$. To simulate moderately informative IE times, we let $T_{Si} = 0.05 * \hat{Y}_i$ where $\hat{Y}_i$ is the predicted event time resulting from a Weibull survival model fit to the PBC data, using the same set of 5 predictors used by L&H. For an uninformative IE, $HDS(t | \Omega(t_0), M^*) = HDS(t | \Omega(t_0), M)$ so that $HDS(t | \Omega(t_0), M^*) / HDS(t | \Omega(t_0), M) = 1$. For illustration, we chose $M = \{age, edema\}$ and $M^* = \{age, edema, I(T_{Si} - t_0), \min(T_{Si}, t_0)\}$ and $t_0 = 1$ year. Figure 1 shows the resulting ratio $HDS(t | \Omega(t_0), M^*) / HDS(t | \Omega(t_0), M)$ averaged across 100 simulations for each setting. As expected, when the IE is uninformative, this estimated ratio is approximately equal to 1. When the IE is mildly or moderately informative, this ratio is greater than 1, with the estimated incremental value varying with t .

While this approach allows for estimation of the incremental value of IE information using HDS, the process of conditioning on individuals that have survived to t_0 may be problematic in terms of inferential procedures if the sample size is small and the events are sparse. In the next section, we describe an alternative approach for confidence interval estimation that may be preferable when sample sizes are small and/or events are sparse.

2 PERTURBATION-RESAMPLING FOR CONFIDENCE INTERVAL ESTIMATION

In finite samples where there may be very few events observed at or close to some timepoints, the proposed HDS estimation and inference procedures may not perform well. Using a simulation study, L&H demonstrate that analytic standard errors provide

approximate nominal coverage rates, and that the coverage rates of percentile bootstrap intervals are closer to nominal levels. For example, in their first simulation study under a Cox model with 25% uninformative censoring, $\widehat{HDS}(t)$ shows little bias with coverage rates using the analytic standard errors ranging from 0.92 to 0.98 when examining t values between 0.025 and 1.0. $\widehat{HDS}^{LC}(t)$ has little bias until $t = 0.60$ and then has moderate bias as the event times are more sparse after this time; coverage rates using the analytic standard errors range from 0.87 to 0.99. When percentile bootstrap confidence intervals are used, the coverage rates improve and range from 0.93 to 0.96.

We propose using a perturbation-resampling method to estimate the variability of the HDS estimates and to construct confidence intervals, similar to the approach proposed in Uno et al. (2013), for estimates of the integrated discrimination improvement with censored time-to-event outcomes. Perturbation-resampling (or ‘wild bootstrapping’) involves generating multiple independent sets of weights from a distribution with mean and variance equal to one. A ‘perturbed’ estimate of the quantity of interest is calculated for each set of weights, and the distribution of these perturbed estimates is used for estimation, as in the standard bootstrap. With this approach, all observations receive some positive weight in each replicated sample. This is especially appealing when data are sparse or outcomes are rare, so that bootstrap resampling may yield samples with no observed events during a time period of interest. Perturbation sampling is an appealing alternative in the case of $HDS(t)$, where inference may be difficult for t in intervals with very few observed event times. Specifically, let $\{V^{(b)} = (V_1^{(b)}, \dots, V_n^{(b)})', b = 1, \dots, B\}$ be $n \times B$ independent copies of a positive random variables V from a known distribution with $E(V) = Var(V) = 1$, such as the standard exponential distribution. Let

$$\widehat{HDS}^{(b)}(t) = \frac{\sum_{i=1}^n V_i^{(b)} \exp \left\{ 2\hat{\beta}^{(b)'} m_i - e^{\hat{\beta}^{(b)'} m_i} \hat{\Lambda}_0^{(b)}(t) \right\} \sum_{i=1}^n V_i^{(b)} \exp \left\{ -e^{\hat{\beta}^{(b)'} m_i} \hat{\Lambda}_0^{(b)}(t) \right\}}{\left[\sum_{i=1}^n V_i^{(b)} \exp \left\{ \hat{\beta}^{(b)'} m_i - e^{\hat{\beta}^{(b)'} m_i} \hat{\Lambda}_0^{(b)}(t) \right\} \right]^2}$$

where $\hat{\beta}^{(b)}$ is the weighted Cox proportional hazards model estimate and $\hat{\Lambda}_0^{(b)}(t)$ is the weighted Breslow estimate, with weights $\{V^{(b)}, b = 1, \dots, B\}$. The distribution of $\{\widehat{HDS}(t) - HDS(t)\}$ can be estimated by the empirical distribution of $\{\widehat{HDS}^{(b)}(t) - \widehat{HDS}(t), b = 1, \dots, B\}$. The perturbation-resampling estimate of the $100(1 - \alpha)\%$ confidence interval for $HDS(t)$ is estimated by calculating the $100(\alpha/2)^{th}$ and $100(1 - \alpha/2)^{th}$ empirical percentiles of $\widehat{HDS}^{(b)}(t)$.

We examine the coverage of perturbation-resampling confidence intervals using L&H’s first simulation setting with 1,000 replicates each and $n = 500$, $n = 300$, or $n = 100$. We use the same data generating function as L&H, with $\lambda(t|M) = \lambda(M) = 0.5 \exp(0.5M_1 + 1.5M_2)$ where M_1 and M_2 are independently distributed Uniform[0, 2]. Results are shown in Table 1 and demonstrate that this alternative approach provides good coverage, even for sample sizes

smaller than 500, though coverage rates were below the nominal 95% level for very small sample sizes ($n = 100$). We attribute this to the relatively few events for small t which resulted in finite sample bias.

While this simulation study is promising, the theoretical properties of the perturbation-resampling approach need to be developed, following arguments similar to those in (Park & Wei, 2003; Cai et al., 2005; Tian et al., 2007). We expect that development of the perturbation-bootstrap approach would be particularly useful for drawing inference for the local-in-time model given the reliance of this method on events within a specific window (bandwidth) of t . It is important to note that if the L&H model assumptions do not hold, one would not expect any of the proposed procedures for confidence interval estimation (the L&H analytic standard error estimator, the bootstrapping procedure or the perturbation-resampling procedure) to provide good coverage for estimates due to expected bias. Future work may be warranted to investigate the robustness of these approaches to departures from model assumptions.

3 OTHER REMARKS

One of the benefits of $HDS(t)$ is the ability to examine changes in the incremental value of predictors over time; some predictors may be especially important early, while others may be more relevant to longer term. By incorporating information about an intermediate events in a landmark model framework, $HDS(t)$ can be used for clinical decision-making such as to plan clinical schedules for disease surveillance and testing after diagnosis.

Lastly, L&H present $HDS(t)$ as an incident risk alternative to the time-varying discrimination slope, $DS(t)$, which is based on cumulative risk. Another difference between these measures is that $DS(t)$ focuses on risk differences, while $HDS(t)$ focuses on risk ratios. It would be useful to have a cumulative risk version of $HDS(t)$ with a more straightforward clinical interpretation.

REFERENCES

- CAI T, TIAN L & WEI LJ (2005). Semiparametric Box-Cox power transformation models for censored survival observations. *Biometrika* 92, 619–632.
- CAREY EJ, ALI AH & LINDOR KD (2015). Primary biliary cirrhosis. *The Lancet* 386, 1565–1575.
- PARAST L, CHENG S-C & CAI T (2011). Incorporating short-term outcome information to predict long-term survival with discrete markers. *Biometrical Journal* 53, 294–307. [PubMed: 21337601]
- PARAST L, CHENG S-C & CAI T (2012). Landmark prediction of long-term survival incorporating short-term event time information. *Journal of the American Statistical Association* 107, 1492–1501. [PubMed: 23293405]
- PARK Y & WEI LJ (2003). Estimating subject-specific survival functions under the accelerated failure time model. *Biometrika* 90, 717–723.
- TIAN L, CAI T, GOETGHEBEUR E & WEI L (2007). Model evaluation based on the sampling distribution of estimated absolute prediction error. *Biometrika* 94, 297–311.
- UNO H, TIAN L, CAI T, KOHANE IS & WEI L (2013). A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Statistics in medicine* 32, 2430–2442. [PubMed: 23037800]
- VAN HOUWELINGEN H & PUTTER H (2011). *Dynamic prediction in clinical survival analysis*. CRC Press.

Figure 1:

Incremental value of simulated intermediate event information using the PBC data as measured by $HDS(t|\Omega(t_0), M^*)/HDS(t|\Omega(t_0), M)$ when the intermediate event is uninformative (gray line), mildly informative (dashed black line), or moderately informative (solid black line); vertical dashed line is drawn at $t_0 = 1$ year

Table 1:

Pointwise coverage results percentile confidence intervals for $\widehat{HDS}(t)$ obtained using perturbation-resampling in the first simulation setting when $n = 500$, $n = 300$, and $n = 100$

time	$n = 500$	$n = 300$	$n = 100$
0.025	0.94	0.93	0.93
0.050	0.94	0.94	0.92
0.100	0.94	0.94	0.92
0.200	0.94	0.94	0.94
0.300	0.95	0.94	0.94
0.400	0.95	0.95	0.94
0.500	0.94	0.95	0.95
0.600	0.94	0.95	0.94
0.700	0.95	0.95	0.95
0.800	0.94	0.95	0.95
0.900	0.94	0.96	0.94
1.000	0.93	0.94	0.94