Review

iMedPub Journals
www.imedpub.com

Biomarkers Journal
ISSN 2472-1646

2021
Vol.7 No.1:80

A Review on Statistical Methods Using Longitudinal Biomarkers for Disease Early Detection

Yongli Han*

Department of Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD 20850, USA

***Corresponding author :** Yongli Han, Department of Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center, Rockville, MD 20850, USA, Email: yongli.han@nih.gov

Received date: January 8, 2021; **Accepted date:** January 22, 2021; **Published date:** January 29, 2021

Citation: Han Y (2021) A Review on Statistical Methods Using Longitudinal Biomarkers for Disease Early Detection. Biomark J Vol.7 No.1:80

Abstract

Longitudinal biomarkers may be predictive for the early detection of clinical outcomes. Tracking longitudinal biomarkers as a way to identify early disease onset may help to reduce mortality from diseases that are more treatable if detected early. This review provides a brief summary of a recent publication on statistical approaches that use longitudinal biomarkers for disease early detection. Comparison of statistical methods and future research directions in general disease early detection using longitudinal biomarkers are presented.

Keywords: Disease early detection; Longitudinal biomarkers; Statistical methods

Introduction

Early detection of diseases before clinical symptoms become present may help to reduce mortality from diseases that are more treatable if detected early [1]. Tracking longitudinal biomarkers in a population may help to detect disease earlier, since changes in serial biomarkers may be indicative for the occurrence of disease [2-4]. An example is ovarian cancer, which usually has no symptoms at its early stage and develops undetected until it has spread within the abdomen and pelvis [5]. Although early-stage ovarian cancer can be treated with a higher success rate [6], the majority of ovarian cancer is diagnosed at late stage, where curative treatment rarely exists [7]. Meanwhile, large randomized trials have not shown a survival benefit for current early detection approaches of ovarian cancer so far [8,9]. Several recently developed statistical approaches, the Risk of Ovarian Cancer Algorithm (ROCA), the Pattern Mixture Model (PMM), and the Shared Random Effects Model (SREM), have demonstrated their capacity for disease early detection [10]. This review provides an overview of these methods through discussing their theories and the comparison in an application to ovarian cancer early detection.

Statistical Methods

To understand the unique patterns of ovarian cancer biomarker trajectories, we considered samples from the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, where the biomarker cancer antigen 125 (CA125) was studied for screening. Trajectories of log transformed CA125 of 50 cases and 50 controls that were randomly selected from the PLCO trial are shown in **Figure 1**. The CA125 profiles for cases may be either flat or stay flat at first and then jump up at some time point during the screening. To the contrary, the control profiles almost always keep flat. The special patterns in the case and control trajectories require advanced statistical modelling strategies for the CA125 behaviors.

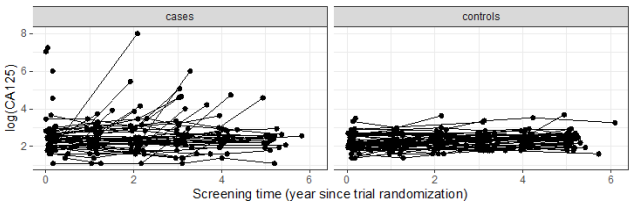

Figure 1: CA125 trajectories of 50 cases and 50 controls that were randomly selected from the PLCO trial.

Risk of Ovarian Cancer Algorithm (ROCA) was explicitly developed for the early detection of ovarian cancer using repeatedly measured CA125 values [11]. In specific, ROCA separately models the longitudinal CA125 trajectories for cases and controls. For controls, a constant mean model of CA125 is assumed with a random intercept term that accounts for subject heterogeneity. As for cases, the averaged CA125 trajectory is presumed to be piecewise linear with a latent subject-specific change point. The risk of having ovarian cancer conditional on CA125 values is then calculated using Bayes' rule. Increased CA125 values would raise suspicion for an undetected tumour and thereby triggers more detailed diagnostic evaluation and intensive follow-up, resulting in possible earlier detection of ovarian cancer and earlier treatments [7]. Yet, several issues of ROCA require attention: (1) the change point is assumed to have a known truncated normal distribution, which may not be reasonable for all study populations; (2) when modeling CA125 profiles, ROCA ignores the effects of screening time and other covariates on marker trajectory change; (3) estimation of the latent change point structure may be unreliable due to measurement sparsity around the change point, which may be caused by long screening interval, e.g., the annual screening design of CA125 in the PLCO trial; and (4) ROCA only obtains an approximated risk, the calculation of which involves

marginalization over the diagnosis time, which is often unknown for a new individual. As ROCA implements the marginalization by borrowing information about the diagnosis time from participants who have already been known as cases, it may result in the loss of prediction accuracy. This is particularly true when the sample size of known cases in a study is relatively small. Nevertheless, extensions of ROCA can be easily made by addressing the above issues.

Pattern Mixture Model (PMM) is a general framework that separately formulates case and control biomarker trajectories [12]. Mathematically, ROCA can be regarded as a special version of the PMM framework. However, there are two key differences between PMM and ROCA in the application of ovarian cancer early detection: (1) instead of using a latent change point structure in the case model, PMM uses natural cubic splines to account for the nonlinear effects of screening time and other baseline covariates on the CA125 profile development; and (2) PMM calculates the exact cancer risk as it directly models CA125 conditional on disease status without marginalization over diagnosis time and hence avoids the loss of prediction accuracy.

Shared Random Effects Model unlike PMM and ROCA, SREM jointly models the combined case and control biomarker trajectories and the binary disease outcome, which are assumed to share the same set of random effects [13]. A linear mixed effects model is assumed for the biomarker profiles while the relation between disease status and random effects is characterized by a probit link function. The joint distribution of the biomarker and disease status can be derived by integrating over the shared random effects, eventually inducing the formula of calculating disease probability.

Discussion

Analyses of the PLCO ovarian cancer data found that the predictive performance of PMM significantly outperformed ROCA and SREM, while the latter two did not have significant difference, under an annual screening setting of the biomarker CA125. All three methods were generally well calibrated. Yet the predictiveness curves showed that PMM and ROCA could place more individuals in the tail areas of the risk distributions than SREM, implying their potential usefulness in the clinical practice of ovarian cancer early detection. Simulation results also showed that those three approaches were generally well calibrated under all data generation scenarios. The predictive accuracy of all methods could be improved under more frequent biomarker screenings, particularly ROCA, as the latent change point structure would be better estimated with more observations around the change point. For practical use, since there is no guarantee that one model would be uniformly better than the other, researchers may apply all three approaches and compare their out-out-sample performances. PMM, ROCA, and SREM can be further applied to other diseases such as prostate cancer, the screening test of which often utilizes the biomarker Prostate-Specific Antigen (PSA), which has similar behaviors as CA125 for ovarian cancer early detection [14].

PMM, ROCA, and SREM are all constructed with a binary outcome and do not fully use the cancer diagnosis time. Therefore, their risk calculations cannot provide absolute risk estimation, that is, the t-year disease-free survival since the last biomarker screening. Landmark approach is an alternative framework when the outcome of interest is time-to-event. This approach would fit a Cox model to individuals who are event-free at a given landmarking time. However, there are several major differences between the landmark approach and PMM, ROCA, and SREM: (1) the landmark approach does not assume structure for longitudinal biomarker but estimates an empirical distribution of the residual survival time based on repeatedly cross-sectional biomarker observations before the land marking time, while the PMM, ROCA, and SREM all use the longitudinal biomarker information for risk prediction; and (2) the landmark method deals with the time-to-event outcome while PMM, ROCA, and SREM all treat the outcome as a dichotomized one, leading to different interpretations of the estimated probabilities [15,16].

Conclusion

Future methodological work may include extensions of PMM and ROCA to the survival model framework and dynamic risk prediction using both longitudinal biomarker values and survival data. One choice for dynamic risk prediction is the Joint Modeling (JM) approach, which jointly models the longitudinal biomarker process and the survival processes through latent shared random effects. However, the model assumptions of JM are often strong and non-verifiable. Another option is the Conditional Modeling (CM) approach, which models the longitudinal biomarker process conditional on the survival time. The CM approach may be sensible, as recent studies have shown that the patterns of longitudinal biomarkers often depend on the terminal event time and the dependence structure may become increasingly evident as the screening time gets close to the terminal event.

Acknowledgement

This study was supported by the Intramural Research Program of the National Cancer Institute, at the National Institutes of Health, United States.

Disclosure Statement

No conflict of interest was reported.

References

1. McIntosh NW, Urban N (2003) A parametric empirical Bayes method for cancer screening using longitudinal observations of a biomarker. Biostatistics 4: 27-40.
2. Drescher CW, Shah C, Thorpe J (2013) Longitudinal screening algorithm that incorporates change over time in CA125 levels identifies ovarian cancer earlier than a single-threshold rule. J Clin Oncol 31: 387-392.
3. Menon U, Ryan A, Kalsi J, et al. (2015) Risk algorithm using serial biomarker measurements doubles the number of screen-detected cancers compared with a single-threshold rule in the United

Kingdom collaborative trial of ovarian cancer screening. J Clin Oncol 33: 2062-2071.

4. Han Y, Liu D (2020) Accounting for random observation time in risk prediction with longitudinal markers: An imputation approach. Stat Methods Med Res 29: 396-412.
5. Matulonis UA, Sood AK, Fallowfield L, Howitt BE, Sehouli J, et al. (2016) Ovarian cancer. Nat Rev Dis Primers 2: 1-22.
6. Clarke-Pearson DL (2009). Screening for ovarian cancer. New England Journal of Medicine, 361(2):170-177.
7. Skates SJ, Greene MH, Buys SS (2017) Early detection of ovarian cancer using the risk of ovarian cancer algorithm with frequent CA125 testing in women at increased familial risk-combined results from two screening trials. Clin Cancer Res 23: 3628-3637.
8. Pinsky PF, Zhu C, Skates SJ (2013). Potential effect of the risk of ovarian cancer algorithm (ROCA) on the mortality outcome of the prostate, lung, colorectal and ovarian (PLCO) trial. Int J Cancer 132: 2127-2133.
9. Wentzensen N (2016) Large ovarian cancer screening trial shows modest mortality reduction, but does not justify population-based ovarian cancer screening. Evid Based Med 21: 159.
10. Han Y, Albert PS, Berg CD, Wentzensen N, Katki HA, et al. (2020) Statistical approaches using longitudinal biomarkers for disease early detection: A comparison of methodologies. Stat Med 39: 4405-4420.
11. Skates SJ, Pauler DK, Jacobs IJ (2001) Screening based on the risk of cancer calculation from Bayesian hierarchical changepoint and mixture models of longitudinal markers. J Am Stat Assoc 96: 429-439.
12. Liu D, Albert PS (2014) Combination of longitudinal biomarkers in predicting binary events. Biostatistics 15: 706-718.
13. Albert PS (2012) A linear mixed model for predicting a binary event from longitudinal data under random effects misspecification. Stat Med 31: 143-154.
14. Lilja H, Ulmert D, Vickers AJ (2008) Prostate-specific antigen and prostate cancer: Prediction, detection and monitoring. Nat Rev Cancer 8: 268-278.
15. Rizopoulos D (2011) Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. Biometrics 67: 819-829.
16. Kong S, Nan B, Kalbfleisch JD, Saran R, Hirth, R (2018) Conditional modelling of longitudinal data with terminal event. J Am Stat Assoc 113: 357-368.