Biostatistical Approaches for Modeling Longitudinal and Event Time Data

Kim-Anh Do
Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030

Introduction
Prostate cancer is the second leading cause of cancer death in United States men, ranking only after lung cancer. In 2002, it is estimated that there will be 189,000 new prostate cancer cases and 30,200 deaths from prostate cancer (1). Early detection of prostate cancer is the focus of today’s clinical research, but its major hindrance is the lack of early symptoms. The introduction of the PSA3 as a marker has allowed diagnosis of prostate cancer at a much earlier stage than was previously possible. The end point of interest may not only be time to onset of prostate cancer but may also include time to recurrence, or survival, which in general may be referred to as event time data or event process data. PSA measurements are influenced by other factors including the amount of normal and malignant prostate tissue, the location and type of cancer, the extent of any existing infection or inflammation of the prostate, the presence of a genetic factor such as the abl gene, or other conditions, such as benign prostatic hyperplasia or prostatitis, among others. Because PSA changes usually occur earlier than a progression or survival end point, serial PSA measurements taken over time are the ideal candidate to aid more accurate prediction of prostate cancer onset and other relevant end points. Clinical investigators have also postulated the existence of different disease patterns manifested in subjective variation of disease progression in certain subsets of men. Therefore, accurate modeling and describing the joint behavior of longitudinal PSA measurements and specific time to event end points of prostate cancer are not only of great clinical and scientific interest but should be the expected requirement for the proper analysis of such important data.

Although sophisticated and novel statistical approaches to handle such data structures abound in statistical and biostatistical journals, they seldom make their way into biomedical journals. On the other hand, important clinical trials are conducted regularly at large and esteemed cancer research centers internationally that result in a wealth of data resources, often analyzed by standard statistical methods, of which the Cox proportional hazards regression model with linear log relative risk functions is a popular tool. In this issue, we welcome and read with interest the article by Verbel et al. (2), who used a flexible time-dependent Cox model with a nonlinear log relative risk function to describe the longitudinal behavior of serial PSA with a survival end point. In addition to the capability of graphing this relation as a smooth nonlinear function, they can also quantify the strength of the association. The authors are longtime researchers in the area of prostate cancer research and modeling (3–5), and their current article brings a breath of fresh air to the analytical methods often presented in Clinical Cancer Research. This commentary supports their effort in bridging the gap between biostatistical/statistical journals and biomedical journals by presenting in summary some of the latest analytical tools developed for the analysis of complicated longitudinal time event data and surrogate markers. We hope that it may stimulate and encourage other contributors to consider using such tools to accurately model and analyze similar data that may arise from other diseases besides prostate cancer.

New Biostatistical Modeling Approaches

The method proposed by Verbel et al. (2) may be susceptible to bias because PSA, modeled as a time-varying predictor, is not continuously observed and must be interpolated. It is perhaps more natural to consider jointly modeling PSA measurements and survival. Furthermore, the nonlinear functional form of the relative risk presented in this article represents the average pattern for the cohort studied. It would be more interesting if an extensive model can capture the subject-specific temporal behavior of PSA that exhibits a characteristic alteration early in the course of prostate cancer and also to disentangle the effects of other established covariates on PSA and survival. Many sophisticated joint modeling approaches for a longitudinal biomarker and an event time process developed in the late 1990s are based on the assumption of common baseline hazard representing a single pattern relating event times to marker trajectories (6–8). A recent development was described by Lin et al. (9), who proposed and applied a latent class joint model of the longitudinal PSA and the prostate cancer onset end point. Their method allows subpopulation structure via latent classes, induces more flexibility to model distinct patterns within each subpopulation, and can be extended to incorporate multiple longitudinal biomarkers. The model specifications include (a) the definition of the probability that a particular subject belongs to a specific latent class described through a multinomial distribution of the class membership vector for this subject, (b) a logit link between the class membership vector and the covariate vector and the associated coefficient vector, and (c) longitudinal PSA measurements for each subpopulation that can be represented by a linear mixed model that can capture common patterns of PSA trajectories within a subpopulation through latent classes while accommodating the variability among subjects in the same class through random effects.

Another new efficient method was described by Wang and Taylor (10) that can resolve the problems of measurement error...
Acknowledgments

I thank Yu Shen for helpful discussion and Marcy Johnson for assistance with literature search.