Contemporary Clinical Trials 36 (2013) 642–650

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

# Bayesian two-step Lasso strategy for biomarker selection in personalized medicine development for time-to-event endpoints

Xuemin Gu [a], Guosheng Yin [b], J. Jack Lee [c,*]

[a] Regeneron Pharmaceuticals, Inc., 110 Allen Road, Basking Ridge, NJ 07920, USA
[b] Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China
[c] Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

ARTICLE INFO

*Article history:*
Received 21 March 2013
Received revised form 16 August 2013
Accepted 16 September 2013
Available online 25 September 2013

*Keywords:*
Adaptive Lasso
Clinical trials
Group Lasso
Proportional hazards model
Targeted therapy design
Variable selection

ABSTRACT

Clinical trial designs for targeted therapy development are progressing toward the goal of personalized medicine. Motivated by the need of ongoing efforts to develop targeted agents for lung cancer patients, we propose a Bayesian two-step Lasso procedure for biomarker selection under the proportional hazards model. We seek to identify the key markers that are either prognostic or predictive with respect to treatment from a large number of biomarkers. In the first step of our two-step strategy, we use the Bayesian group Lasso to identify the important marker groups, wherein each group contains the main effect of a single marker and its interactions with treatments. Applying a loose selection criterion in the first step, the goal of first step is to screen out unimportant biomarkers. In the second step, we zoom in to select the individual markers and interactions between markers and treatments in order to identify prognostic or predictive markers using the Bayesian adaptive Lasso. Our strategy takes a full Bayesian approach and is built upon rapid advancement of Lasso methodologies with variable selection. The proposed method is generally applicable to the development of targeted therapies in clinical trials. Our simulation study demonstrates the good performance of the two-step Lasso: Important biomarkers can typically be selected with high probabilities, and unimportant markers can be effectively eliminated from the model.



## 1. Introduction

Cytotoxic chemotherapies continue to be the primary form of treatment for cancer. The treatment effects of cytotoxic agents come from their ability to eradicate rapidly dividing cancer cells. However, such detrimental effects are not specific to cancer cells, as rapidly dividing normal cells (e.g., hair or bone marrow) are often harmed by cytotoxic agents as well, and thus often results in side effects with varying severity. Historically, both the development and application of chemotherapy in treating cancer have been largely based on a specific cancer diagnosis, which is determined by the location and microscopic appearance (histology) of the tumor. Thus, patients diagnosed with cancer of the same classification are typically given the same treatment. Research has shown, however, that patients with similar tumor histologies may respond differently to the same chemotherapy. Hence, the administration of chemotherapy guided by a traditional tumor diagnosis may expose patients to excessive toxicity and result in unwanted side effects.

Unprecedented advances in life science, mainly during the last two decades, have revolutionized the landscape of cancer drug development via an exciting concept known as "personalized medicine". The development of new molecularly targeted therapy is a major thrust to seize the promise held by personalized medicine. The new way for cancer drug development involves identifying specific regulators that play key roles in various processes of cancer biology, and developing molecularly targeted

* Corresponding author.
*E-mail addresses:* xuemin.gu@regeneron.com (X. Gu), gyin@hku.hk (G. Yin), jjlee@mdanderson.org (J.J. Lee).

1551-7144/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cct.2013.09.009

agents for these regulators to block signaling events associated with the growth of tumors. Unlike traditional approaches, personalized medicine uses novel diagnoses to screen for patients who are most likely to benefit from specific treatments based on an association between the molecular profiles of patients and the targeted effect of a specific therapy. This approach then assigns treatments that are individually tailored to patients according to their own molecular profiles.

As recognized in the Clinical Path Initiative, a program created by the U.S. Food and Drug Administration (FDA), one important component of targeted agent development is the identification and validation of biomarkers as molecular targets for patient screening and clinical endpoint evaluation. Current technological capabilities in genomics and proteomics allow researchers to quickly collect a large amount of biomarker information from patients in a cost-effective fashion. Therefore, the key issue in the design of clinical trials for personalized medicine is the ability to identify important and meaningful predictors from a pool of many possible variables.

Based on functions in diagnosis and treatment selection for cancer patients, biomarkers can be roughly classified into two categories: prognostic markers and predictive markers. Prognostic markers reflect a healthy status or a disease stage of a patient; they are associated with disease outcomes regardless of the treatment. One obvious prognostic biomarker is age. Older ages usually imply shorter survival times on all treatments. In prostate cancer, a common prognostic biomarker is the prostate-specific antigen (PSA), for which a higher value of PSA reflecting a larger tumor burden and, consequently, poor prognosis of a patient. On the other hand, biomarkers that can predict differential treatment efficacy in different marker groups are called predictive markers. For example, a high level of human epidermal growth factor receptor 2 (HER-2) is a predictive marker for trastuzumab, a targeted breast cancer therapy approved by the FDA. In the clinical trials for targeted therapy development, one primary goal is centered around the identification and validation of predictive and prognostic markers. In the linear model setting, treatment effects are usually characterized by a linear combination of treatment main effects, marker main effects, and marker–treatment interactions. In this case, a non-zero marker main effect represents a prognostic marker and a non-zero marker–treatment interaction signifies a predictive marker.

Our research is motivated by one of the first biopsy-based and biomarker-integrated clinical trials for targeted agent development at the MD Anderson Cancer Center. The trial is referred to as BATTLE, which stands for "Biomarker-based Approaches of Targeted Therapy for Lung Cancer Elimination" as described in [10,16]. One of the main aims of the BATTLE trial is to establish a program for clinical trials with targeted therapy development. The BATTLE trial also seeks to identify molecular features in tumor tissues that correlate with tumor response and to discover new signaling pathways to be tested in future trials. Building from the success of the original BATTLE trial, several follow-up trials are being planned at the MD Anderson Cancer Center with primary goals of validating the findings in the BATTLE trial and identifying biomarkers associated with treatment effects of novel combinations of targeted therapies. The selection of variables from numerous biomarkers is an important aspect of these new trials for targeted therapy development.

To embrace the demand of emerging targeted agent trials, we propose a two-step variable selection strategy for the time-to-event endpoint to identify important biomarkers, as the selected biomarkers can be subsequently used in the adaptive randomization procedure to assign more patients with better treatments based on patients' marker profiles. Hence, the variable selection must be accurate and robust, meaning that the selected biomarkers should be able to provide good prediction and the selection should be stable against variation in the data. The least absolute shrinkage and selection operator (Lasso) [14] and its various extensions are suited to this purpose, as Lasso can handle variable selection and parameter estimation simultaneously. Lasso is a natural choice of the statistical approach to targeted agent development, for which marker identification and treatment effect estimation are equally important. To better incorporate the variable selection process into the Bayesian adaptive design framework of our ongoing trial, we implement the Bayesian Lasso, which simplifies the selection of the tuning parameter and takes into consideration the uncertainty of variable selection. We only focus on the variable selection part of our Bayesian adaptive design here. Our marker selection strategy consists of two sequential steps: Step 1 uses the Bayesian group Lasso to screen for biomarkers with either prognostic or predictive values for grouped variables (with each biomarker group as a selection unit); and step 2 applies the Bayesian adaptive Lasso for refined variable selection among the biomarkers identified in the first step. Our simulation study demonstrates that this Bayesian two-step Lasso strategy outperforms the usual one-step Lasso variable selection methods.

It often occurs in oncology trials that many participants have already failed at least one prior treatment, and the experimental new drug may be their only hope for effective disease control. One incentive for participation in a clinical trial is the potential of providing effective treatments to patients within the trial. To find effective treatments for each patient, biomarkers that can differentiate treatment effects among patients need to be identified. Lasso-type methods have been extensively developed in the context of variable selection, which offers a suitable tool for modeling covariate effects in targeted agent development. Although only the variable selection strategy is presented here, the combination of our method with Bayesian adaptive randomization design can help us to achieve the goal of treating patients better, as we continue to learn from accumulating data to identify important biomarkers for treatment selection and progressively optimize patient allocation based on biomarker information.

The remainder of this paper is organized as follows. In Section 2, we provide an introduction to Lasso and its Bayesian implementation under the Cox proportional hazards model, and propose a Bayesian two-step Lasso strategy motivated by the need for biomarker identification for targeted agent development. In Section 3, we examine the performance of the Bayesian two-step Lasso method through simulation studies in terms of identifying both prognostic and predictive markers. We conclude with some discussions in Section 4.

## 2. Bayesian two-step Lasso strategy

### 2.1. Bayesian Cox's proportional hazards model

In clinical trials for targeted agent development, the time-to-event (TTE) outcomes, such as progression-free survival (PFS) or overall survival (OS), are typically the clinically