Principal Investigator

First Name: Lei
Last Name: Zhang
Degree: PhD
Primary Affiliation: Bristol Medical School, University of Bristol
E-mail: leizhang8821@gmail.com
Phone number: +44 (0) 117 928 7237
Address:
CANYNGE HALL, 39 WHATLEY ROAD
City: BRISTOL
State or Province: Bristol
Zip or Postal Code: BS8 2PS
Country: United Kingdom

General Information

Key Personnel (in addition to PI):

Are external grants or funds being used to support this research?: No external grants or funds are being used to support this research.
How did you learn about the YODA Project?: Internet Search

Conflict of Interest

https://yoda.yale.edu/system/files/yoda_project_coi_form_for_data_requestors_2019_5.pdf

Certification

Certification: All information is complete; I (PI) am responsible for the research; data will not be used to support litigious/commercial aims.
Data Use Agreement Training: As the Principal Investigator of this study, I certify that I have completed the YODA Project Data Use Agreement Training

1. NCT00270283 - CR006076 (I88-009) - A Double-Blind, Placebo-Controlled Study With Open-Label Follow-up to Determine the Safety and Efficacy of Subcutaneous Doses of r-HuEPO in AIDS Patients With Anemia Induced by Their Disease and AZT Therapy

What type of data are you looking for?: Individual Participant-Level Data, which includes Full CSR and all supporting documentation

Research Proposal

Project Title

A new fixed-effects approach for validation of a longitudinally measured surrogate biomarker for a time-to-event endpoint

Narrative Summary:

The studies on validation of the surrogate biomarkers in medical fields entail the data on repeat measurements of the biomarkers and the elapsed time to an event. The measures of biomarkers may encounter some measurement errors which need to be adjusted using random-effects (RE) model before measuring its association with the time-to-event endpoint. However, the endogeneity bias in the RE model may bias the inference on the effectiveness of biomarker as surrogate for the disease state or the failure time. This project will propose a new fixed-effects (FE) model which aims to produce more trustworthy estimates of the biomarkers for validating its surrogacy than the RE model.

Scientific Abstract:

Background: surrogate biomarker relating to disease state, the failure time or whether a new treatment has some beneficial effects on the time to a certain clinical event has always been prominent in clinical trials research. The traditional approach using the random-effects (RE) model to monitor the latent process of the biomarker over the follow up remains a challenge because of the potential provision for biased results induced by the unobserved patient-level confounders. The fixed-effects (FE) approach provides another way to describe the trajectory of the biomarker within the study period. The FE model places fewer restrictions on assumptions made of the independence between covariates and cluster effects of the outcome in the model, and this flexibility can be desirable for providing more precise parameters in analyses of validating the surrogate biomarkers than the RE model.
Objective: We propose a new FE model which provides a better estimator and inference for validating the surrogate biomarkers than the traditional FE model and RE model.
Study design: We will apply the new FE, traditional FE and RE model to analyze the trial data to validate the haemoglobin as a surrogacy in assessing whether the new r-HuEPO treatment has beneficial effects on reducing the transfusion risk for HIV patients with anemia.
Participants: All 102 patients in the trial we are requesting.
Main Outcome Measures: Patients’ survival time.
Statistical Analysis: We will apply the new FE model to analyze the data and compare the inferences to results from the old FE and RE models.

Brief Project Background and Statement of Project Significance:

In many clinical trial studies, considerable attention is paid to evaluation of the longitudinal measurements of surrogate biomarkers with disease state, the failure time or whether a new treatment has some beneficial effects on the time to a certain clinical event. The standard statistical approach suggests to first adjust the measurement error, within-patient variability and the between-patient biologic variability for the observed sequence of the biomarkers using the growth curve model (e.g., Tsiatis et al., 1995).The association between biomarker and event can be then modeled using Cox model by including the adjusted biomarker measures at baseline, the time point when the measures reach the peak or bottom during the follow up and survival time as covariates.
Despite the recognition of the value of the growth curve model when modelling its trajectory over time in this setting, we should also be aware of the disadvantages when applying the model to the real data (Skrondal & Rabe-Hesketh, 2004). The challenge arises because, as one of the RE models, the growth curve model requires that all variables must be independent of the cluster effects (exogeneity of covariates with cluster effects); this requirement cannot be met in most analyses in particular the studies in this setting. Some previous literature (e.g., Renard et al., 2003) has shown that the biomarker measurements tend to be strongly associated with the covariates (e.g., time points of the follow up). These correlations are induced by some unobserved patient-level variables (e.g., patients’ underlying health conditions or biologic responses to different therapy), which could result in significantly biased estimates of the random effects and other coefficients(Bell & Jones, 2015). In addition, there is a concern that the RE approach would fail to converge or produce unnecessarily large standard errors for random effects for data from small clinical trials due to the limitation of sample size (Snijders & Bosker, 2012).
Alternatively, the FE model provides another useful way to adjust the biomarker measures. Because of its non-exogeneity assumptions on cluster effects and covariates, the FE model has great significance in eliminating the endogeneity bias and in advising small standard errors for the coefficients in the model (Rabe-Hesketh & Skrondal, 2008). However, the traditional FE model is not widely used in this setting. The reason is that, though it is important to further control the patient-level non-varying variable (e.g., dummy for treatment and control group) in the model to describe the different trajectories of the biomarkers across different groups, the traditional FE model would not allow us to do so because adding the patient-level non-varying variable in the model will cause a rank-deficiency failure in its OLS estimation (Lockwood & McCaffrey, 2007). In this project, we propose a new FE model which resolves the rank-deficiency failure in the old FE model. This new FE model will provide a better estimator for

making inference of validation for surrogate biomarkers, especially for analyses with small sample size.

Specific Aims of the Project:

Specific Aims:
1) Propose a new FE model which allows to control the patient-level non-varying variable in the model.
2) Demonstrate the theorem as to why the OLS estimators in the new FE model can resolve the rank-deficiency failure in the traditional FE approach and avoid the endogeneity bias induced by the unobserved patient-level variables in the RE model.
3) Apply the new FE, traditional FE models and the growth curve model to analyze the requested data as a case study.
Objectives:
1) By demonstrating the mathematical theorem, we will show that the new FE model can provide a better estimator than the RE and old FE models for a broad range of analyses in terms of avoiding rank-deficiency failure and endogeneity bias.
2) By analyzing the study data from a small clinical trial, we will show the advantage of using the new FE model to adjust endogeneity bias in research on validating biomarker's surrogacy, and its great feature in analyses with small sample size where the RE model could fail to converge.
Hypotheses:
1)The new FE model provides a good alternative for research on validating the biomarker’s surrogacy. It can be extremely useful for analyses with small sample size where the RE and old FE models would fail to produce precise parameters.

What is the purpose of the analysis being proposed? Please select all that apply.
Confirm or validate previously conducted research on treatment effectiveness
Develop or refine statistical methods

Research Methods

Data Source and Inclusion/Exclusion Criteria to be used to define the patient sample for your study:

We will evaluate out methodology by analyzing data from all patients who participated in the randomized trial and whose data are available at YODA. Since our objective is to evaluate a new statistical method, we will not apply any inclusion or exclusion criteria for selecting patients.

Main Outcome Measure and how it will be categorized/defined for your study:

We plan to analyze the data from the clinical trial we are requesting as a case study. As this project aims at comparing different approaches in validating haemoglobin as a surrogacy in assessing whether the new r-HuEPO treatment has beneficial effects on reducing transfusion risk for AIDS patients with anemia, the main outcome measures will be patients’ survival time in the trial (The primary endpoint is a haematocrit of 38-40% or 12 weeks).

Main Predictor/Independent Variable and how it will be categorized/defined for your study:

The main predictor are longitudinal measurements of haemoglobin within the study period.

Other Variables of Interest that will be used in your analysis and how they will be categorized/defined for your study:

We are also interested in including the dummy indicator for the treatment/control group in the model, which aims to describe the difference in the trajectories of haemoglobin between the two groups.

Statistical Analysis Plan:

In this study we will first propose the new FE model with the illustration of its OLS estimator for patient effect of the biomarkers and its standard error. How this new FE model resolves the rank-deficiency failure in traditional FE model will be also demonstrated. We will also state the theorem as to why the OLS estimator in this new FE approach can avoid the endogeneity bias induced by the unobserved patient-level variables. Some simulation

studies will be conducted to show how the nature of endogeneity biases in estimates of the random slope and random intercept from a growth curve model depends on the magnitude and direction of the correlations between covariates and the omitted patient-level variables represented by the random effects.
In the case study we plan to analyse the trial data as follows:
1) The sequence of the haemoglobin measurements from baseline to end point will be modelled by the new FE model, traditional FE model and the growth curve model respectively. The nonlinear term of time points (e.g., polynomial or spline function) will be included in all models to capture the trajectory of the haemoglobin. In RE approach we will allow the effects of time points to vary across patients by adding the random slopes on the time t. The dummy indicator for treatment/control group will be adjusted in the new FE model and growth curve model since the patients in the two groups may have different trajectories because of the different treatments (the traditional FE model would not allow us to include this variable in the model). The OLS estimates from the both FE approaches and the Empirical Bayes (EB) estimates from the RE approach will be produced to predict the adjusted values of the biomarker at specific time points (e.g., baseline, 4 weeks of follow up and survival time t) for each patient.
2) We further assume that the patients' survival time in the trial depends only on the underlying pattern of the haemoglobin decline or rise, not the biomarker observations themselves (De Gruttola & Tu, 1994). The underlying pattern of the biomarker within the study period will be captured by the adjusted values of the biomarker at baseline, the time point when the measures reach the peak or bottom (e.g., at 4 weeks or 8 weeks) during the follow up (or use the random slope of the time point directly instead in RE approach), and survival time t . To validate the haemoglobin as surrogacy in assessing whether the new r-HuEPO treatment has had beneficial effect, we will evaluate the difference in the treatment effects of r-HuEPO estimated in cox models before and after controlling these adjusted measures of haemoglobin. The ideal surrogate biomarker for the new treatment would explain most of the difference in hazard ratio between the treatment and control group after controlling the adjusted biomarker values. To evaluate the performances of the new FE, traditional FE and growth curve model on providing trustworthy adjusted haemoglobin values for validating its surrogacy, we will compare the results of the proportions of the treatment effect of r-HuEPO on survival time explained by the adjusted biomarker produced by these three approaches.
Software Used:
STATA

Project Timeline:

We plan to finish all the analyses within 6 months after we gain access to the data. We plan to have an initial manuscript ready by the end of 12 months.

Dissemination Plan:

We will publish the method and the analyses (as a case study) in a statistical journal, such as “Statistics in Medicine”, “Statistical Methods in Medical Research” or “Journal of American Statistical Association”.

Bibliography:

Bell, A., & Jones, K. (2015). Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. Political Science Research and Methods, 3(01), 133-153. Bian, Y., Breiger, R., Galaskiewicz, J., & Davis, D. (2005).
De Gruttola, V., & Tu, X. M. (1994). Modelling progression of CD4-lymphocyte count and its relationship to survival time. Biometrics, 1003-1014. -252.
Lockwood, J. R., & McCaffrey, D. F. (2007). Controlling for individual heterogeneity in longitudinal models, with applications to student achievement. Electronic Journal of Statistics, 1, 223.
Rabe-Hesketh, S., & Skrondal, A. (2008). Multilevel and longitudinal modelling using Stata (2nd ed.): STATA press.
Renard, D., Geys, H., Molenberghs, G., Burzykowski, T., Buyse, M., Vangeneugden, T., & Bijnens, L. (2003). Validation of a longitudinally measured surrogate marker for a time-to-event endpoint. Journal of Applied Statistics, 30(2), 235-247.
Skrondal, A., & Rabe-Hesketh, S. (2004). Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models: CRC Press.
Snijders, T. A. B. ,& Bosker, RJ. (2012). Multilevel analysis: An introduction to basic and advanced multilevel modeling (2 ed.): SAGE.
Tsiatis, A. A., Degruttola, V., & Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. Journal of the American statistical association, 90(429), 27-37.