Principal Investigator

First Name: Gary
Last Name: Rosner
Degree: ScD
Primary Affiliation: Johns Hopkins University
E-mail: grosner1@jhmi.edu
Phone number: 410-955-1198
Address: 550 N. Broadway
Suite 1103
City: Baltimore
State or Province: Maryland
Zip or Postal Code: 21205
Country: United States

General Information

Key Personnel (in addition to PI):
First Name: Chenguang
Last name: Wang
Degree: Ph.D.
Primary Affiliation: Johns Hopkins University
SCOPUS ID: 35977003800

First Name: Shutong
Last name: Du
Degree: BS
Primary Affiliation: Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health
SCOPUS ID:

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?: Colleague

Conflict of Interest

https://yoda.yale.edu/system/files/signed_conflictofinterestform.pdf
https://yoda.yale.edu/system/files/yoda_project_coi_form_for_data_requestors_2016_cg.pdf
https://yoda.yale.edu/system/files/yoda_project_coi_form_for_data_requestors_du.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. NCT00211133 - CR004414 (EPO-INT-76) DOUBLE BLIND - A Double-blind, Randomized, Placebo-controlled Study to Evaluate the Impact of Maintaining Hemoglobin Using Eprex (Epoetin Alfa) in Metastatic Breast Carcinoma Subjects Receiving Chemotherapy
2. NCT00270127 - CR005917 (EPO-INT-10) - Double-Blind, Placebo-Controlled Study to Assess the Effect of

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

Research Proposal

Project Title

Bayesian Methods for Comparative Effectiveness Research

Narrative Summary:

Comparative effectiveness research entails combining relevant data from disparate sources. Information about the effect and effectiveness of a treatment for a disease, such as cancer or heart disease, may come from randomized clinical trials, early phase non-randomized studies, and hospital or payer databases. One needs to account for inherent differences between the studies, however, because heterogeneities may bias the inference. This project will develop a flexible inferential model using Bayesian nonparametric methods to characterize prior uncertainty and allow for borrowing strength when appropriate, while also accommodating heterogeneities between data sources.

Scientific Abstract:

Background: Comparative effectiveness research (CER) focuses on evaluations of treatment strategies in real-world settings. A treatment's effectiveness information may come from randomized clinical trials (RCTs), non-randomized studies, and hospital or payer databases. The Bayesian inferential paradigm provides a structured way to combine information from disparate sources within the framework of probability theory. Bayesian nonparametric (BNP) models place fewer restrictions on assumed mathematical forms of underlying probability distributions, and this flexibility may improve inference concerning key parameters, such as treatment effects.

Objective: We propose to extend a published BNP model, the hierarchical Dirichlet process mixture model, in a CER analysis inferring a treatment's effectiveness from multiple randomized clinical trials and possibly observational data. We will also evaluate this model with propensity score adjustment.

Study Design: We will apply our BNP model to analyze patient-level data from 8 RCTs evaluating Epoetin alfa. We may also perturb the raw data from some RCTs to mimic non-RCT data to evaluate the model.

Participants: All patients in the 8 studies for whom we have the same information, e.g., hemoglobin change, transfusion rate.

Main Outcome Measures: The likely outcome measures for our analyses will be hemoglobin and transfusion rate (if all 8 RCTs include these outcomes).

Statistical Analysis: We will apply the BNP hierarchical model to analyze the data and compare inferences to results from less flexible meta-analytic models.

Brief Project Background and Statement of Project Significance:

Comparative effectiveness research often combines relevant data from disparate sources. Data about a treatment's effectiveness may come from randomized clinical trials (RCTs), non-randomized studies, and hospital or payer databases. Appropriate statistical models for combining RCT data with information from patients undergoing the same treatment strategy in clinical practices should help comparative effectiveness inferences. The Bayesian inferential paradigm provides a structured way to combine information from disparate sources within the framework of probability theory. The ultimate question is how best to treat the next patient who enters the clinic. In Bayesian statistics, this question relates to the predictive distribution. Given the current data, how likely are key outcomes if one chooses to treat this patient with the new treatment or with a different one? The Bayesian inferential framework allows one to answer this and related questions.

Heterogeneity between sources of data may exist, and one needs to account for inherent differences for valid inference. For example, RCT patients satisfy strict entry criteria. Also, outcomes in the RCT setting with its prescribed supportive care may well differ from outcomes among eligible patients at the same institution who decline to participate or from outcomes in community hospital settings whose patients may not have access to the same level of care. One may model these differences as covariates when these covariates are available, but unmeasured factors may also affect the treatment outcome.

Regression approaches infer the effects of covariates as shifts of expected values on some scale (e.g., linear or logistic). Differences in the patient populations may alter the underlying distributions beyond a simple shift in the location and lead to between-study differences for which regression methods are inadequate. Bayesian nonparametric (BNP) methods provide a more flexible approach to inference. One approach is based on mixtures of simple distributions. For example, mixtures of normal distributions can characterize almost any distribution, be it skewed, multimodal, etc.

We have proposed flexible Bayesian analytic methods based on mixtures for meta-analysis. The model decomposes the distribution of parameters into a common distribution shared by all studies and study-specific distributions. The final distribution is a weighted average of the common and study-specific distributions. The common and study-specific distributions may be finite mixture models (Lopes, Müller, and Rosner, 2003) or infinite Dirichlet process (DP) mixtures (Müller, Quintana and Rosner, 2004). The models also allow nonparametric regression on covariates.

Earlier papers did not fully evaluate the hierarchical DP mixture model with a large number of studies. We propose applying the model to analyze RCT data. Aside from analyses of study data, we will also evaluate how robust the inference is by perturbing data from some of the RCTs and analyzing the resulting data. We feel that BNP models will allow one to evaluate effectiveness by including more relevant data sources than other models allow.

Specific Aims of the Project:

Specific Aims:
1) Extend a Bayesian Dirichlet process mixture model for time-to-event and count data.
2) Demonstrate the behavior of the hierarchical Bayesian Dirichlet process mixture model in the presence of varying degrees of between-study heterogeneity.
3) Create a publicly available R package for carrying out meta-analyses using a Bayesian Dirichlet process mixture model.
4) Evaluate the Bayesian Dirichlet process mixture model with propensity score adjustment.

Objectives:
• Evaluate the usefulness of a Bayesian Dirichlet process mixture model for meta-analysis when studies are not exchangeable.
• Include propensity score adjustment in the Bayesian Dirichlet process mixture model and evaluate the resulting inference, compared to a non-mixture model analysis.

Hypotheses:
• The Bayesian Dirichlet process mixture model will provide better inference than commonly used hierarchical models when studies are not fully exchangeable.
• The Bayesian Dirichlet process mixture model will be able to adjust for imbalances better than other regression-type approaches, since it does not assume a linear relationship on some scale.

What is the purpose of the analysis being proposed? Please select all that apply.
Other

Development of new 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 trials evaluating erythropoietin and whose data are available at YODA. Since our objective is to evaluate a 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:

The main outcome measures for this project will be the outcomes of the randomized clinical trials we are requesting. The requested trials of erythropoietin do not appear to have the same primary endpoints, necessitating us to find a meaningful clinical endpoint that each data set contains. The two most likely outcomes will be hemoglobin and transfusion rate (number of transfusions per unit time, such as one or more months).

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

We will evaluate known covariates that affect the outcome, such as baseline hemoglobin. Additionally, we will use all covariates that are common to the randomized trials for the purpose of developing propensity scores.

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

There are no other variables of interest.

Statistical Analysis Plan:

We will analyze the data using the hierarchical Dirichlet process mixture (HDPM) model described in the Background section. We will compare the results to inferences using a Bayesian hierarchical model that makes more assumptions about the distribution of model parameters across studies. We will also evaluate the robustness of the HDPM model by perturbing the data in some studies to mimic registries and carrying out an analysis of all studies. We will also evaluate the use of propensity score adjustment in the context of the HDPM model.

We may have data from fewer than 8 studies to use for our analysis. Our goal is to evaluate a Bayesian nonparametric (NP) model for use in comparative effectiveness research (CER), not to carry out a meta-analysis. We plan to use data from these trials primarily to create an environment similar to what one might encounter in CER. We have not found randomized clinical trial (RCT) data and a publicly available registry relating to the same treatment. We want to use these trials' data as comprising a hypothetical registry plus RCTs' data. E.g., we may accomplish the CER scenario by treating data from the active EPO alpha arm of one or two RCTs as registry data, ignoring each trial's comparator arms. By combining these "registry" data with data from the remaining 6 RCTs, we can imitate a situation that could arise in practice. The actual clinical trial data allow for a more realistic evaluation than would simulation data. Of course, the number of RCTs with usable (for our purpose) outcome data may turn out to be fewer than eight. As long as we have at least 4 RCTs, we feel we can carry out our intended analyses. We also propose evaluating the use of propensity scores within a Bayesian NP framework for CER. Whereas Bayesian regression tree (BART) methodology is a useful tool for carrying out propensity score-based analysis, we want to consider situations in which one may not have full exchangeability of the data sets one is analyzing. We feel that the particular HDPM model we will use will provide better inferences in situations for which one can assume partial exchangeability.

The setting we describe in the proposal is not a meta-analysis of RCTs. Instead, we want to consider situations that might arise in CER. We feel that the appropriate comparator for CER is a Bayesian hierarchical regression-type model rather than standard approaches for meta-analysis.

Our model is a hierarchical version of a Dirichlet process (DP) mixture model. A DP places a probability distribution on the space of probability distributions. The posterior distribution with a DP prior is almost surely discrete, which can be awkward if one wants to characterize a continuous distribution. If, however, one assumes a DP prior for the distribution of a probability distribution's parameter(s), then the posterior distribution will be a mixture of these distributions centered at discrete locations, with the posterior DP as the mixing measure. E.g., if one assumes that the mean of a normal distribution has a DP prior, then the posterior will be a DP mixture of normals. The locations of this mixture will come from the DP posterior for the mean. In this way, the posterior can be thought of as a mixture of normals, even though the posterior is a DP mixture.

We will apply the propensity score adjustment approach to all studies and infer the treatment's effect from all available data. The propensity score adjustment will be based on our proposed HDPM model approach. Roughly speaking, we will conduct a NP regression of the clinical outcomes on the propensity scores in a CER setting with the Bayesian NP method described in the proposal and in our response to the third comment. The Bayesian NP prior leads to a NP regression after integrating over the mixtures. Furthermore, our analyses will utilize the strength of the proposed HDPM models to allow information sharing among multiple randomized and/or observational studies.

Project Timeline:

We plan to have the R program ready in 3 months. The initial analyses will be finished 7 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 "Biometrics" or "Biostatistics."

We will publish a separate manuscript describing incorporation of propensity score adjustment. We will publish this paper in a statistical journal, such as "Biometrics," or in a journal with more of a clinical trials focus, such as "Clinical Trials."

We will provide an R package for implementing our model as part of the publication.

Bibliography:

Lopes HF, Müller P, Rosner GL. Bayesian meta-analysis for longitudinal data models using multivariate mixture priors. Biometrics. Mar 2003;59(1):66-75.
Müller P, Quintana F, Rosner GL. A Method for Combining Inferences Across Related NonParametric Bayesian Models. Journal of the Royal Statistical Society Series B-Statistical Methodology. 2004;66(3):735-749.

Supplementary Material:

https://yoda.yale.edu/sites/default/files/lopesmuellerrosner_bayesianmetaanalysisforlongitudinaldatamodelsusingmultivariatemixturepriors.pdf
https://yoda.yale.edu/sites/default/files/muellerquintanarosnerjrssb2004.pdf