Principal Investigator

First Name: Chunhua
Last Name: Weng
Degree: PhD
Primary Affiliation: Columbia University
E-mail: cw2384@columbia.edu
Phone number: 212-305-3317
Address: 622 W 168 Street PH-20

City: New York
State or Province: NY
Zip or Postal Code: 10032
Country: USA
SCOPUS ID: 8979750700

2015-0649

General Information

Key Personnel (in addition to PI): **First Name:** Anando
Last name: Sen
Degree: PhD
Primary Affiliation: Columbia University

Are external grants or funds being used to support this research?: No external grants or funds are being used to support this research.

yoda_project_coi_form_for_data_requestors_2015.pdf
yoda_project_coi_form_for_data_requestors_2015-signed_sen.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

Associated Trial(s): NCT00036374 - A Randomized, Double-Blind Trial of Anti-TNF Chimeric Monoclonal Antibody (Infliximab) in Combination With Methotrexate for the Treatment of Patients With Polyarticular Juvenile Rheumatoid Arthritis
NCT00036439 - A Randomized, Placebo-controlled, Double-blind Trial to Evaluate the Safety and Efficacy of Infliximab in Patients With Active Ulcerative Colitis
NCT00096655 - A Randomized, Placebo-controlled, Double-blind Trial to Evaluate the Safety and Efficacy of Infliximab in Patients With Active Ulcerative Colitis
NCT00207675 - A Randomized, Multicenter, Open-label Study to Evaluate the Safety and Efficacy of Anti-TNF a Chimeric Monoclonal Antibody (Infliximab, REMICADE) in Pediatric Subjects With Moderate to Severe CROHN'S Disease

NCT00094458 - Multicenter, Randomized, Double-Blind, Active Controlled Trial Comparing REMICADE® (infliximab) and REMICADE plus Azathioprine to Azathioprine in the Treatment of Patients with Crohn's Disease Naive to both Immunomodulators and Biologic
NCT00119756 - A Randomized, Crossover Study to Evaluate the Overall Safety and Tolerability of Paliperidone Palmitate Injected in the Deltoid or Gluteus Muscle in Patients With Schizophrenia
NCT00210548 - A Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Dose-Response Study to Evaluate the Efficacy and Safety of 3 Fixed Doses (50 mg eq., 100 mg eq., and 150 mg eq.) of Paliperidone Palmitate in Subjects With Schizophrenia
NCT00101634 - A Randomized, Double-blind, Placebo-controlled, Parallel-group, Dose-response Study to Evaluate the Efficacy and Safety of 3 Fixed Doses (25 mg eq, 50 mg eq, and 100 mg eq) of Paliperidone Palmitate in Patients With Schizophrenia
NCT00076115 - Research on the Effectiveness of Risperidone in Bipolar Disorder in Adolescents and Children (REACH): A Double-Blind, Randomized, Placebo-Controlled Study of the Efficacy and Safety of Risperidone for the Treatment of Acute Mania in Bipola
NCT00132678 - A Randomized, Double-blind, Placebo-controlled Study to Explore the Efficacy and Safety of Risperidone Long-acting Intramuscular Injectable in the Prevention of Mood Episodes in Bipolar 1 Disorder, With Open-label Extension
NCT00094926 - A Prospective, Randomized, Double-blind, Placebo-controlled Study of the Effectiveness and Safety of RISPERDAL CONSTA Augmentation in Adult Patients With Frequently-relapsing Bipolar Disorder
NCT00207662 - ACCENT I - A Randomized, Double-blind, Placebo-controlled Trial of Anti-TNFa Chimeric Monoclonal Antibody (Infliximab, Remicade) in the Long-term Treatment of Patients With Moderately to Severely Active Crohn's Disease
NCT00037674 - A Randomized, Double-Blind, Multicenter, Placebo-Controlled 12-Week Study of the Safety and Efficacy of Two Doses of Topiramate for the Treatment of Acute Manic or Mixed Episodes in Patients With Bipolar I Disorder With an Optional Open-La
NCT00035230 - A Randomized, Double-Blind, Multicenter, Placebo-Controlled 12-Week Study of the Safety and Efficacy of Topiramate in Patients With Acute Manic or Mixed Episodes of Bipolar I Disorder With an Optional Open-Label Extension
A Randomized, Double-Blind, Multicenter, Placebo-Controlled, 21-Day Study of the Safety and Efficacy of Topiramate for the Treatment of Acute Manic or Mixed Episodes in Subjects With Bipolar I Disorder With an Optional Open-Label Extension
NCT00202865 - Evaluation of Low Dose Infliximab in Ankylosing Spondylitis (CANDLE)
NCT00077714 - A Randomized, Double-blind, Placebo- and Active-controlled, Parallel-group, Dose-response Study to Evaluate the Efficacy and Safety of 2 Fixed Dosages of Paliperidone Extended Release Tablets and Olanzapine, With Open-label Extension, in t
NCT00083668 - A Randomized, Double-blind, Placebo- and Active-controlled, Parallel-group, Dose-response Study to Evaluate the Efficacy and Safety of 3 Fixed Dosages of Paliperidone Extended Release (ER) Tablets and Olanzapine, With Open-label Extension,
NCT00074477 - A Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Efficacy and Safety of 50 and 100 Mg-eq of Paliperidone Palmitate in Patients With Schizophrenia
NCT00078039 - Trial Evaluating Three Fixed Dosages of Paliperidone Extended-Release (ER) Tablets and Olanzapine in the Treatment of Patients With Schizophrenia

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

Research Proposal

Project Title

Population Representativeness of Clinical Trial Study Samples

Narrative Summary:
Population representativeness of clinical trial study samples directly influences the generalizability of clinical trial results. We aim to develop a multivariate representativeness measure to quantify the population representativeness of the study samples of each clinical trial. We will compare the clinical trial study samples to real-world patients using this measure.

Scientific Abstract:
Background:
The lack of population representativeness in clinical trial study samples has been a persistent problem that has been observed across disease domains. Numerous studies have shown the differences between real-world patients and clinical trial study samples. This problem is significant in that the lack of population representativeness can compromise the generalizability of clinical trial results and lead to unforeseen adverse events among the real-world patients. It is imperative to improve the transparency of the population representativeness of clinical trial samples using systematic analyses.

Objective:
Our study aims to develop a quantitative metric that is able to quantify the population representativeness of each trial using multiple study traits while accounting for their dependencies and relevancies.

Study Design:
We will design and validate the population representative metric by comparing patient-level clinical trial sample to real-world target populations made available by electronic health records. For each trial, we will identify the real-world target population in our institution's clinical data warehouse, which contains 4.5 million patients' 20 years of data. We will profile the study samples and the target population for each trial and measure the representativeness of study sample of the target population.

Participants:
all enrolled patients in the trials.

Main outcome measures:
a score of population representativeness of each trial

Statistical analysis:
our metric for representativeness is statistically-based

Brief Project Background and Statement of Project Significance:
In every study, investigators need to extrapolate study results from the Study Sample, which comprises all the individuals enrolled in a study, to the Target Population, which comprises all individuals for whom a treatment may be considered for its intended purpose. The Sample population is recruited from and is a subset of a Study Population, which is defined by the eligibility criteria and comprises all individuals who would be eligible to enroll in that study. Thus the population representativeness of the Study Population is a function of the eligibility criteria and directly influences the generalizability of the study results. Ideally, the definition of the study population should be optimized to reflect the target population. Since it is impossible to accurately profile the Target Population A, we can approximate A by profiling all the members of A who have EHR data, which we call Target Population with EHRs.

Our core hypothesis is that the more accurately clinical investigators can characterize the target population as defined by EHR data -- the better informed and prepared they will be able to define eligibility criteria to construct the study population to appropriately represent the true Target Population. This research will help cultivate a new culture of clarity and transparency about eligibility criteria, in which investigators are expected and supported to explain and justify the rationale for their eligibility criteria choices. Such transparency will enable more effective measures to reduce clinical research biases and improve research reliability.

This research promises to improve the transparency of the population representativeness of clinical research eligibility criteria and hence reduce biases and improve research reliability, and enable early iterative in silico estimates of feasibility and population representativeness

Specific Aims of the Project:
The specific aims of this project are
(a) developing a multi-feature representativenss metric
(b) measuring the population representativness of the study samples of the included 123 trials in the Yoda database

What is the purpose of the analysis being proposed? Please select all that apply. New research question to examine treatment safety

Participant-level data meta-analysis
Participant-level data meta-analysis uses only data from YODA Project

Research Methods

Data Source and Inclusion/Exclusion Criteria to be used to define the patient sample for your study:
Data sources:
1. all study samples in the Yoda database: we will include both clinical and demographic characteristics for profiling the study samples in clinical trials. For example, in our 2014 paper, we compared the distributions of A1c and age between the real-world diabetes patients and the study populations (derived from eligibility criteria) of 1761 Type 2 diabetes trials. We will look at the inclusiveness of individual traits as well as the combination of all the traits used in clinical trials. We prefer to use individual-level data instead of summary statistics because our methodology for quantifying population representativeness involves the identification of the fraction of the diseased population who are part of the trial. Using individual-level data will provide more precise results than summary statistics and allows to uncover overly restrictive eligibility criteria and provide this feedback to trial designers. Morever, existing summary statistics inlcudes only mean, max, or min values; however, we need to use other measures that are more appopriate than mean etc.

2. EHR-based target population characteristics from our institutional clinical data warehouse

Main Outcome Measure and how it will be categorized/defined for your study:
1. population representativeness of each clinical trial based on the combination of multiple eligibility criteria/study traits
2. population representativeness of each eligibiltiy criterion/each trait

Main Predictor/Independent Variable and how it will be categorized/defined for your study:
None

Other Variables of Interest that will be used in your analysis and how they will be categorized/defined for your study:
None

Statistical Analysis Plan:
We will use a quantitative metric to analyze population representativeness. All traits will be weighted by using EHR data to model the dependencies between multiple traits (also accounting for feature weights). The metric has several of the expected mathematical properties which will be verified through simulation-based evaluations. We are measuring the "likelihood" of a clinical trial's findings being applied to patients who were not a part of the trial based on (a) eligibility criteria (b) actual trial patients. We will use all features (both quantitative and categorical) that are a part of the eligibility criteria of the study. The individual patient data will be used to evaluate what fraction of the diseased population is represented in the trials.

We will be happy to provide a copy of our manuscript with details for our quantitative metric for population representativeness upon publication. We have used other public data sources to develop this metric and will be interested in applying this metric on the Yoda database.

Project Timeline:
we expect to use 1 year to complete this project:

months 1-3: data collection
months 4-6: study sample and target population profiling
months 4-9: population represetnativess analysis and evaluation
months 6-12: publication

Dissemination Plan:
we will publish and present our methods and findings in peer-reviewed mainstream informatics journals and conferences.

Bibliography:

Weng C, Optimizing Clinical Research Participant Selection with Informatics, Trends in Pharmacological Sciences 36 (2015), Cell Press, pp. 706-709. PDF

Ma H, Weng C, Identification of Questionable Exclusion Criteria in Mental Disorder Clinical Trials Using a Medical Encyclopedia, 2016 Pacific Symposium on Biocomputing, in press. (Travel Awardee)

He Z, Chandar P, Ryan PB, Weng C, Simulation-based Evaluation of the Generalizability Index for Study Traits, AMIA Annu Fall Symp, 2015, in press. PDF (Distinguished Paper Award)

He Z, Wang S, Borhanian E, Weng C, Assessing the Collective Population Representativeness of Related Type 2 Diabetes Trials by Combining Multiple Public Data Resources, Proc of MedInfo'2015, Sao Paulo, Brazil, 19-23 August, accepted. PDF

He Z, Carini S, Sim I, Weng C, Visual Aggregate Analysis of Eligibility Features in Clinical Trials, Journal of Biomedical Informatics, 2015 Jan 20. pii: S1532-0464(15)00007-6. PMID: 25615940 PDF

He Z, Carini S, Hao T, Sim I, Weng C, A Method for Analyzing Commonalities in Clinical Trial Target Populations, Proc of AMIA 2014 Fall Symp, Nov 15-19, 2014, Washington DC, 1777-1786. PMID: 25954450

Weng C, Li Y, Ryan P, Zhang Y, Gao J, Liu F, Bigger JT, Hripcsak G, A Distribution-based Method for Assessing The Differences between Clinical Trial Target Populations and Patient Populations in Electronic Health Records, Applied Clinical Informatics, Vol. 5: Issue 2 2014, 463-479. PMID: 25024761 PDF

Weng C, Yaman A, Lin K, He Z, Trend and Network Analysis of Common Eligibility Features for Cancer Trials in ClinicalTrials.gov, Smart Health Lecture Notes in Computer Science Volume 8549, 2014, pp 130-141

* Hao T, Rusanov A, Boland MR, Weng C, Clustering Clinical Trials with Similar Eligibility Criteria Features, J Biomed Inform, 2014 Feb 1. pii: S1532-0464(14)00011-2. PMID: 24496068