NETWORK MODELS OF CLINICAL TRIAL DATA

Marco D. Sorani[1], Geoffrey T. Manley[1],
J. Claude Hemphill[1] and Sergio E. Baranzini[2]
[1]*Department of Neurological Surgery and* [2]*Department of Neurology*
University of California, San Francisco
San Francisco, CA 94110, USA
Emails: soranim@pharmacy.ucsf.edu, manleyg@neurosurg.ucsf.edu,
chemphill@sfgh.ucsf.edu, sebaran@cgl.ucsf.edu

Networks can reveal complex relationships in large data sets. We obtained data for clinical trials from a public-access registry. We assembled "node-edge-node" triplets to construct networks of diseases and interventions for visualization and analysis using Cytoscape software. We downloaded trials for Stroke, Multiple Sclerosis and Brain Injuries. There were significant differences in trial design and composition across indications..

1. INTRODUCTION

Network topologies have been shown to share features across diverse fields of study.[1-3] Network models have often been applied to biological systems, but they are now also being implemented to investigate clinical questions. With the growth of public-access registries, it is becoming possible to apply network models to clinical trial data.

In this study, we apply network models to describe the multi-level structure of clinical trials obtained from the clinicaltrials.gov registry (http://www.clinicaltrials.gov).

2. MULTI-LEVEL NETWORKS

For a set of over 6,000 trials, we downloaded multiple parameters and performed several transformations on the data. We obtained National Clinical Trial ID, Condition, Intervention, Sponsor, Start Date and Completion Date for each trial.

Where possible, we standardized conditions to Medical Subject Headings (MeSH) (http://www.ncbi.nlm.nih.gov/mesh) nomenclature. Clinicaltrials.gov uses the term "Condition" for both their pre-defined diseases and the free-form text submitted by investigators for each trial. Intervention data are also submitted by investigators as free-form text. There may be dozens of free-form text Conditions, some of which reflect true sub-phenotypes[4] and some of which are cases of inconsistent nomenclature.

We constructed networks for visualization using Cytoscape software (http://www.cytoscape.org). Once we constructed the networks, we analyzed their topologies visually and quantitatively. In this study, networks are displayed using the yFiles (http://www.yworks.com) Organic layout. We also identified hubs and clusters in the networks.

We constructed multi-level networks to analyze trial Conditions and Sponsors. We defined Condition levels as the clinicaltrials.gov diseases and the free-form text Conditions. We defined two Sponsor levels as sets of Sponsors and individual Sponsors. Networks had a large, primary connected component subdivided into more or less distinct clusters which correspond to higher levels of aggregation, such as disease category.

3. NETWORK STATISTICS

We downloaded trials for Stroke, Multiple Sclerosis and Brain Injuries (1027, 517, and 469 trials, respectively). There were significant differences across indications in recruitment status, gender composition, age groups, and phases ($p<0.001$) by chi-squared tests. Differences in intervention type ($p<0.001$) demonstrated more frequent drug interventions in MS (80%) than stroke (42%) but more frequent device interventions in stroke (21%) than MS (5%). Numbers of edges in trial-condition (135,820; 61,587; 15,957) and trial-intervention (2466; 2341; 519) networks were rank-correlated with trial counts. Fewer connected

components suggested stronger connectivity in MS and brain injury versus stroke (21 and 24 versus 40).

4. DISCUSSION

In this study, we present a network-based analysis of clinical trial data. Using a large set of data from clinical trials obtained from the clinicaltrials.gov registry, we examine the topological parameters, network hubs, and clinical features of clinical trial conditions, interventions and sponsors. We propose solutions to defining nomenclature for constructing clinical trial networks. We also find that different disease types demonstrate divergent network topologies.

There were significant differences in trial design and composition across indications. Given the approximate 2:1:1 ratios of trials, the ratios of edges in trial-condition and trial-intervention networks suggest fewer shared study conditions and interventions among brain injury trials, particularly compared to MS trials. Network models may be useful tools to provide insight into high-level relationships among clinical trials.

There are several potential future directions for this work. We chose to take a disease-centric approach by focusing on relationships between trials in neurological diseases. Alternative approaches could include expanding scope, by looking at all diseases. To further integrate multiple levels of detail into network models of clinical trials, it would be useful to have patient-level data, but this may be difficult to obtain. Other issues to investigate include scale-dependence[5] and network vulnerability[6] if the findings of a "hub" trial are called into question.

In conclusion, we have presented a network-based analysis of public clinical trials data. We defined a large set of trials in neurological conditions using data from clinicaltrials.gov. We analyzed multi-level models that integrated levels of granularity of trial conditions, interventions, and sponsors. We also performed visual and topological evaluations. We highlight opportunities to make trial nomenclature more computable, and we describe divergent network topologies in different disease types.

References

1. Clauset A, Moore C, Newman ME. Hierarchical structure and the prediction of missing links in networks. *Nature* 2008; **453**:98-101.
2. Butts CT. Revisiting the foundations of network analysis. *Science* 2009; **325**:414-416.
3. Tero A, Takagi S, Saigusa T, Ito K, Bebber DP, Fricker MD, Yumiki K, Kobayashi R, Nakagaki T. Rules for biologically inspired adaptive network design. *Science* 2010; **327**:439-442.
4. Miller G. Neuroscience. New guidelines aim to improve studies of traumatic brain injury. *Science* 2010; **328**:297.
5. Gotelli NJ, Graves G, Rahbek C. Macroecological signals of species interactions in the Danish avifauna. *Proc Natl Acad Sci USA* 2010; **107**, 5030-5035.
6. Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S. Catastrophic cascade of failures in interdependent networks. *Nature* 2010; **464**:1025-1028.