Geographic and Network Analysis of Oncology Trials: Portfolio Assessment of ClinicalTrials.gov

Steven K Cheng, PhD[1,2], Bradford R Hirsch, MD MBA[3,4], Robert M Califf, MD[5], Asba Tasneem, PhD[4], John Horton, MS[4], Kevin A Schulman, MD[4], Amy P Abernethy, MD[3,4], David M Dilts, PhD MBA[1,2]

[1]Center for Management Research in Healthcare; [2]Knight Cancer Institute, Oregon Health & Science University - Portland Oregon
[3]Duke Cancer Institute; [4]Duke Clinical Research Institute; [5]Duke Translational Research Institute - Durham North Carolina

Financial support for the AACT database was provided by grant U19FD003800 from the U.S. Food and Drug Administration awarded to Duke University for the Clinical Trials Transformation Initiative

Introduction

Increasing complexity and specificity of cancer trials necessitates collaboration across the clinical trials network to optimize research efforts. The relationship among clinical trial sites and networks, as observed in ClinicalTrials.gov, provides unique insights to improve coordination and accrual.

Methodology

5,971 interventional clinical trials registered between 2007 and 2010 were aggregated by trial and geographic location. Clinical trials were regrouped into clinical specialties; the team used the National Library of Medicine's (NLM) MeSH thesaurus (2010 version) and submitted conditions [1]. Clinical trials for the ten most represented cancer types by number of sites were assessed using network graph theory and geographic analysis comparing distribution of trials across metropolitan statistical areas (MSAs). This analysis was stratified by mid-phase (Phase I/II, II) and late-phase (Phase II/III, III) trials.

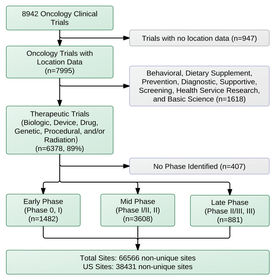

Figure 1: Identification of clinical trials and sites

Incidence & Number of Clinical Trials

Number of trial locations by disease types were categorized by MSAs based upon zip code designation. Number of trial locations across MSAs were compared against cancer incidence by specific cancer types using Pearson Correlations (figure 2a-c). Incidence levels are based on the CDC Wonder - United States Cancer Statistics by Metropolitan Statistical Areas [2].

Network Analysis

Network density represents the degree of dyadic interconnection between sites studying similar cancer types. For each clinical trial in the identified sample, individual trial locations were designated by a node - all sites within a trial were linked into a cluster. Spring force layout [4] was applied to each node after all trials nodes were mapped (figures 3a-c). The algorithm attempts to place each node as far as possible from other nodes; while, at the same time, connectors tie nodes together. The weight of a connector denotes the similar number of share trials in the portfolio, the stronger the weight of each connector, the closer nodes will be pulled together (attractive force). Density is the calculated across the entire network indicating the number of actual connections between nodes relative to the total number of possible connections (table 1).

Geographical Analysis

Nodes are mapped to location for geographical representation (figures 4a-c). Nodes with a high number of trials conducted in the zip code are color-coded. Red nodes contain more than 15 trials of the specific disease type and phase and are highly interconnected across the entire network; yellow nodes contain more than 15 trials of the specific disease type and phase, but are not highly inter-connected across the network. Green nodes contain less than 15 trials of the specific disease type and phase.

[1] Califf, RM; Zarin, DA; Kramer, JM; Sherman, RE; Aberle, LH; Tasneem, A. "Characteristics of Clinical Trials Registered in ClinicalTrials.gov, 2007-2010. Journal of American Medical Association.
[2]Tasneem A, Aberle L, Ananth H, et al. The database for Aggregate Analysis of ClinicalTrials.gov (AACT) and subsequent regrouping by clinical specialty. PLoS One. 2012;7(3):e33677
[3] National Program of Cancer Registries Early Release Cancer Statistics: 1999-2009, WONDER On-line Database. United States Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; 2012.
[4] Kamada, Tomihisa; Kawai, Satoru. "An algorithm for drawing general undirected graphs". Information processing letters, 1989.

Phase III Trials

Figure 2a-c: Incidence and Number of Clinical Trials by MSAs

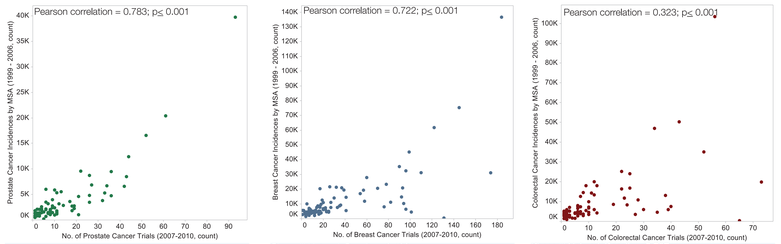

Figure 3a-c: Network Diagrams by Cancer Types

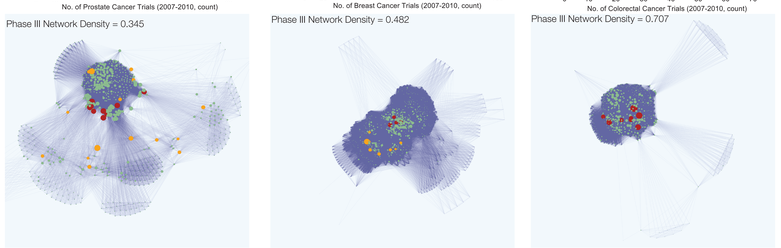

Figure 4a-c: Geographical Map of Trials

size of the node reflects the number of phase III trials being conducted within a zip code
red nodes = greater than 15 phase III trials within a single zip code & network centrality ≥ 0.75
yellow nodes = greater than 15 phase III trials within a single zip code & network centrality < 0.75
green nodes = zip codes with less than 15 phase III trials

Table 1: Correlation of Trial Availability and Incidence Compared with Network Density Across Cancer Types, by Phase

Top Ten Cancer Types (by Number of Sites)	Correlation between Incidence and trial availability (by MSA)	Mid-Phase (Phase I/II, II)			Late-Phase (Phase II/III, III)		
		No. of Sites	No. of Trials	Network Density	No. of Sites	No. of Trials	Network Density
Lymphoma	0.890	2745	319	0.257	1470	50	0.298
Myeloma	0.798	670	155	0.080	1537	44	0.473
AML	0.790	1343	143	0.315	805	26	0.559
Prostate	0.783	1167	205	0.081	3457	56	0.345
Breast	0.722	3891	359	0.242	4842	113	0.482
Lung	0.649	5124	401	0.175	5080	79	0.548
CNS	0.592	1421	172	0.399	842	23	0.567
Pancreas	0.583	1275	135	0.142	424	24	0.424
Liver	0.565	911	125	0.395	1524	37	0.659
Colorectal	0.323	2213	223	0.173	1793	62	0.707
TOTAL	**0.797**	**30401**	**3608**		**31844**	**748**	

Results

66,566 clinical trial sites were identified across the sample - 59.6% of sites were located in the United States. There are a greater number of mid-phase trials than late-phase trials across all cancer types; but the number of total clinical trial sites by phase is dependent on cancer types.

Network analysis can be used as a surrogate measurement to assess collaboration across sites and nodes. Within the United States, the correlation between number of trials within geographical regions and local cancer incidence rates is highly significant (correlation = 0.797, p≤ 0.001) (table 1). Overall, mid-phase trial networks are less dense with sparser interconnections among sites than late-phase trials. For late-phase trials, cancer types with higher network density have an inverse correlation with geographic distribution by incidence rates (correlation = -0.777, p= 0.008) (figure 4). For example, late-phase lymphoma trials have a low network density (density = 0.298) and have a high correlation across geographic-specific incidence rates (correlation = 0.890). Mid-phase trials do not show a similar trend.

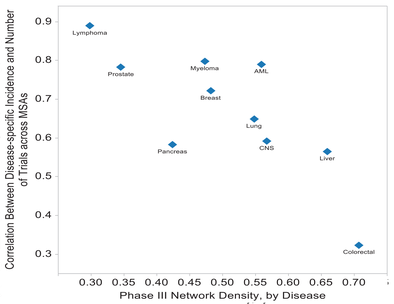

Figure 4: Correlation of Network Density and Local Incidence Rates Across 10 Cancer Types

Detailed investigation of nodal characteristics by geographic regions for late-phase trials reveal that areas with a high volume of highly interconnected trial across network (red nodes) have multiple sites in same region that open similar trials. Nodes with high-volume but non-connected trials across the network (yellow nodes) have a larger proportion of single and smaller multi-institutional trials.

Conclusions

ClinicalTrials.gov database provides a unique insight into the implicit and explicit networks across disease sites. Network structures vary significantly depending upon phase and cancer type. The relationship between the distribution of CT and location of patients for accrual can be envisaged through geographic and network analysis of clinical trials registry data. Late-phase high-density networks should strive to diversify trial locations to better meet regional incidence rates; late phase low-density networks should strive to coordinate clinical research across the United States.

Next Steps & Limitations

- Characterization of nodes based on site capabilities and demographics to conduct comparative network capabilities
- Incorporation of site-level outcomes data including per-site enrollment to better understand the impact of networks of clinical trials relative to total clinical trial outcomes
- Comparison of oncology trial networks against other disease types including cardiovascular and mental health