Clinical Trials.Gov: A Topical Analyses

Vibha Anand, PhD[1], Amos Cahan, MD[1], Soumya Ghosh, PhD[1]
[1]IBM T.J. Watson Research Center, Cambridge, MA.

Abstract

ClinicalTrials.gov was established as a web-based registry for clinical trials of human participants in 2000. Mandatory registration started in 2008. Given more than a decade of registered trials, it's important to understand the "topic" areas and their evolution over time from this resource. This information may help in identifying current knowledge gaps. We use dynamic topic model (DTM) methods to discover topics and their evolution over last 17 years. Our model suggests that there are disease or organ specific trials such as 'Cardiovascular disorders', Heart & Brain conditions', or 'Breast & Prostate cancer' as well as trials registered for general health. General health trials are less likely to be FDA regulated, but both health and pain management, as well as surgical, heart, and brain trials have upward trend in recent years while advanced cancer trials have downward trended. Our model derives unique insights from metadata associated with each topic area.

Introduction

ClinicalTrials.gov (www.clinicaltrials.gov) is a web-based registry and results database for clinical studies of human participants that are conducted around the world. It was created as a result of the Food and Drug Administration (FDA) Modernization Act (FDAMA) of 1997 and the web site was publicly made available in 2000. The registry and site are maintained by the National Library of Medicine (NLM) and serves as an important resource for both patient families and health care professionals as well as researchers alike. The site provides information on publicly and privately supported clinical trials on a wide range of diseases and conditions conducted globally. Typically, at the start of a clinical trial, the sponsor or the principal investigator of the study provides the initial information for registration and during the study period, this information may be updated as needed.

The trials registry mostly contains records on clinical trials or interventional studies (requiring human subjects assigned to medical interventions based on a protocol) but some records also exist for observational studies and expanded access trials. Additionally, those clinical trial types that are not required by the law may not be registered on the site, but over time voluntary registration has increased as well as policies and laws have been enacted for mandatory registration. For example, the FDA Amendments Act of 2007 (FDAAA) made registration and reporting of study results mandatory including information on study participants, summary outcomes, including adverse events mandatory for certain types of trials. Section 801 of the FDAAA also established certain penalties in failing to do so. As a result of the policies, the ClinicalTrials.gov was also augmented with a results database which became publicly available in September 2008. The recently published Final Rule was developed by the Health and Human Services to clarify the requirements for reporting summary study results on ClinicalTrials.gov with the purpose of improving researchers and sponsors compliance.[1] There are over 224,000 studies from all across the world that are registered on the ClinicalTrials.gov website, with 23,000 of them listing summary results. [1]

Thus the ClinicalTrials.gov registry and results database serve as an invaluable health science resource for practitioners, researchers and patient community alike. However, for any further meta-analyses of trials or to pursue any comparative effectiveness research using ClinicalTrials.gov records, the dataset not only needs to be downloaded but also converted to a format where it can be used for further analyses. [2] This exercise is not trivial due to the many variations on how the study participant characteristics, results and outcomes are defined and reported in the database.[2] Furthermore, the meta-analyses techniques themselves are extremely time consuming to perform by hand and may not cover the entire breadth of coverage. For example, they may be often limited to the practicality of analyses at hand and many study topic areas may remain unexplored. While time-consuming and extremely difficult to produce, meta-analyses do shed light on the overall effect (or not) of an intervention and help address experimental bias in clinical trials by pooling data from multiple studies,[3,4] They also influence clinical guidelines for use in practice and spur new research directions.[5,6] In fact the Cochrane Collaboration, a volunteer-based organization leverages a volunteer workforce of 37,000 people for publishing systematic reviews.[7] Automating meta-analyses of clinical trials have also been attempted in the past [8] but their scope has been limited.

Given the importance of meta-analyses and that the ClinicalTrials.gov registry and database was established over a decade ago with mandatory reporting of certain trial types starting in as early as 2008, it is important to know the

overarching study "topic" areas in these registered trials. This information may help in identifying current knowledge gaps before embarking on meta-analyses or comparative effectiveness research. Additionally, it will be useful to know if certain study areas are more or less sponsored and by what kind of sponsors; the characteristics of the study participants and how these topic areas have evolved over time. This information may also help drive comparative effectiveness research to evaluate different treatment options from prior clinical trials and inform judicious allocation of funding resources in future.

More recently, tools have been developed to navigate the registry site. Using query tools one can search study records based on certain criteria (disease or condition). Records meeting the criteria can also be downloaded for further analyses. However, most of these tools provide summary information of the location and types of studies (interventional vs. observational) and study results by year. They may also describe overall trends using charts or maps, however none of the existing tools provide a thematic view of the registry database.

Topic models are probabilistic algorithms that can annotate documents with thematic information. [9] By analyzing word co-occurrence patterns, they discover underlying themes or "topics" from collections of documents. For example, analyzing news articles might result in the discovery of themes spanning "weather", "financial", "political", "sports" and "current affairs". At a high level, these methods can discover how these themes are connected to each other and how they change over time. In this study, using dynamic topic modeling (DTM) methods[10], we analyze information in the ClinicalTrials.gov registry to discover high level thematic areas or "topics" from clinical trials registered in the last decade or more, their trends over time and information such as study participant characteristics or study center and sponsor types associated with each of them.

Methods

We fit a topic model to the ClinicalTrials.gov registry records. A topic model is a generative model, whereby the assumption is that the topics are defined even before the documents are "generated" from them. The task at hand then is to learn a statistical model from a collection of documents or "corpus". This model reflects that collectively the documents exhibit multiple topics, and that each document may exhibit one or more topics, however in different proportions. Additionally, if ordering of the documents in the corpus is important, a dynamic topic model can be used to elicit longitudinal changes in topics. We first describe the dataset from the ClinicalTrials.gov registry used in this study and then describe the DTM modeling method [10] used to automatically discover the hidden topics and their trends in the dataset. To derive useful insights from the learned model, such as the most likely disease or condition in a given topic, we match the metadata fields (Table 1) from ClinicalTrials.gov records to the model's output (matrix of per document topic probability proportion) based on the original document title index. We empirically choose a threshold of 90th percentile as cutoff for documents of interest. The results from documents of interest by topic are described under the heading "Insights from Metadata" in the Results section.

Dataset:

From the ClinicalTrials.gov web site, we initiated an "Advanced Search" (August 12, 2016) without providing any filtering criteria such as study terms or dates. The search resulted in 218,628 records which were downloaded as a tab separated Excel file. This output file contains information for each trial registered in the ClinicalTrials.gov registry database. Each record has a study index (NCT Number), a "Title", and "First Received" date among other fields. Of these, we used "Title" and "First Received" as inputs for learning a DTM (Please see model section below). We assumed that the combination of these fields were unique in our entire dataset and we chose all studies between 2000 and 2016 for our analyses, excluding anything prior to 2000. This resulted in 218,618 records. Other fields included in the output file are metadata fields for each trial such as "Conditions" (a disease or condition associated with the trial), "Gender" (of study participants), or "Phase" of the trial. Please see Table 1 for details of all fields in the output file and their characteristics. However, the Advanced Search tool does not provide complete information for every trial, for example "Study Sponsor" or "Is_FDA_Regulated" or "Is_Section_801" fields are not present in the output of the query tool. To seek these, we downloaded the September 2015 version of the database for Aggregate Analysis of ClinicalTrials.gov (AACT). [11] AACT is a companion of the website and contains such metadata information. The downloaded version of the AACT includes studies that were registered and publicly released before 25 September 2015. To get the additional metadata fields for each trial, the output file from the "Advanced Search" tool were joined with the AACT records (based on NCT Number). This resulted in a total of 181,690 records that were available for our final metadata analyses for each topic.

Table 1: Metadata in ClinicalTrials.gov registry used in this study (n=218,618)

	Variables	Description	Count
1.	Title	Unique title record of trial	217,047
2.	Conditions	Disease condition(s) for trial	79,752
3.	First Received	Date in years	17
4.	Gender	Both, Female, Male	
5.	Overall status	Status of the trial – Completed, Recruiting, Terminated, Active, Not Recruiting, Withdrawn	
6.	Study Types	Interventional, Observational, Expanded Access	
7.	Is Section 801	Yes, No	
8.	Is FDA Regulated	Yes, No	

Dynamic Topic Model (DTM)

Topic models are an example of generative latent variable models. Such models assume that the observed data is produced by a generative process governed by hidden random variables. The generative process defines a joint probability distribution over the observed and hidden random variables. The aim then is to learn conditional probability distributions or posterior distributions over the hidden variables, given the observed variables and the joint probability distribution. A classic example of a topic model is Latent Dirichlet Allocation (LDA). LDA formally defines a "topic" as a distribution over a fixed vocabulary of words in a document collection. The observed variables in the model are the words in each document. LDA assumes that a document corpus expresses a set of K topics, where K is empirically chosen. Words in a document are then generated according to the following generative process. A word in a document is drawn from one of the K topics, where the selected topic (topic assignment for the word) is chosen according to a per-document distribution over topics.

The goal then is to automatically discover the hidden structure, i.e. "topics", per-document topic distributions, and per document per word topic assignments in the collection. Please see Figure 1 for details. Thus in a LDA model each document exhibits all K topics, but in different proportions. The inference problem in LDA is to learn the hidden structure given the observed words in each document in the collection. LDA makes certain assumptions, first, the number of topics to be discovered are known, second, the ordering of words ("bag of words") in the documents or for that matter ordering of documents in the collection does not matter. However, for large corpuses that run over years, such as from the ClinicalTrials.gov registry, this assumption is not appropriate. Modeling the temporal order of documents in these collections is important for understanding topic evolution over time Therefore, one approach to this problem is to use dynamic topic models, which combine topic models with state-space models to model longitudinally evolving topics. In DTM methods, a "topic" is defined as a *sequence of distributions* over words rather than a single distribution of words. Thus, if a collection is arranged by years, each year is a "time slice" in the model and documents of each slice form a component topic model where the topics associated with each slice evolve from the topics from the previous slice. This change facilitates tracking topic evolution or thematic change in the collection over time. Unfortunately, posterior inference in the dynamic topic model is intractable and one has to resort to approximate inference techniques. Here we employ structured variational inference techniques to infer an approximation to the posterior [3].

DTM from ClinicalTrials.gov registry:

From the ClinicalTrial.gov dataset (Table 1), we used the "Title" variable for the document collection and "First Received" (year in the date) for defining the number of "time slices" in our DTM model. Based on preliminary models, we empirically chose 15 topics. There are 17 distinct years from the First Received field in the registry records used for these analyses, and these correspond to 17 time slices, i.e. 2000 – 2016 in the model. We used an open source tool the Natural Language Tool Kit (NLTK, www.nltk.org), and a free python library for topic modeling - Gensim (http://pydoc.net/Python/gensim/0.11.1/gensim.models.wrappers.dtmmodel/). NLTK is used to build the dictionary and corpus from the document collection and Gensim wrapper is used to set the seed parameters for learning DTM model. We learned four separate models with different seeds. These models were evaluated for performance using the expected lower bound metric [12], and the best performing model was chosen for final analyses in this study. The model produces per-document topic proportions for each year in 2000 to 2016.

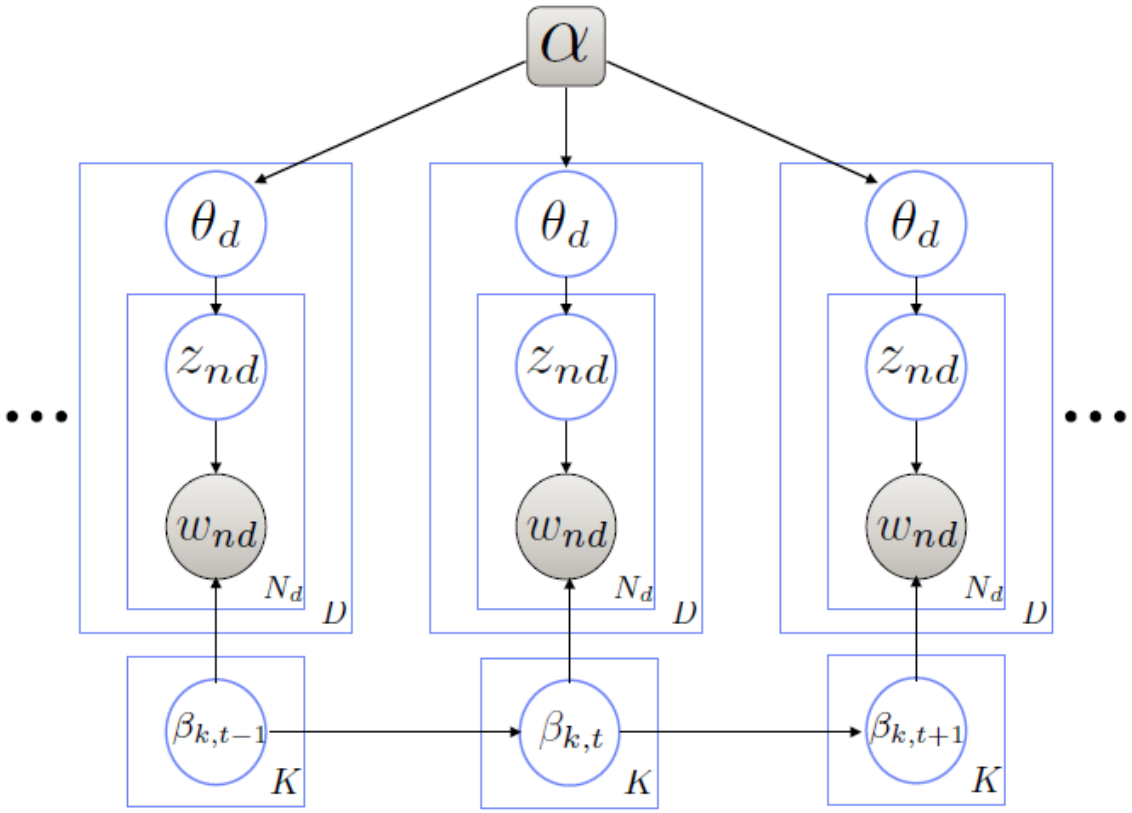

Figure 1: Graphical model summarizing the conditional dependencies assumed by the dynamic topic model (DTM). The model assumes the following generative procedure. A document d at time t is generated by sampling a document specific distribution over topics $\theta_d \mid \alpha \sim \text{Dir}(\alpha)$ from a Dirichlet distribution. Words are generated by first selecting a topic $z_{dn} \mid \theta_d \sim \text{Categorical}(\theta_d)$ from the document specific topic distribution and then selecting a word from that topic $w_{dn} \mid z_{dn}, \beta_{k,t} \sim \text{Categorical}(\mathcal{S}(\beta_{z_{dn},t}))$, where $\mathcal{S}(v) = \frac{\exp(v)}{\sum_{v'} \exp(v')}$ is the softmax function that maps $v \in \mathbf{R}^M$ to the M dimensional probability simplex ($\sum_{j=1}^{M} v_j = 1$). Longitudinal effects are modeled by allowing the unnormalized topic representations $\beta_{k,t}$ to evolve over time, $\beta_{k,t} \sim \mathcal{N}(\beta_{k,t-1}, \sigma^2 I_M)$, where σ^2 is a noise parameter that controls how much topics can change between two time steps, I_M is a $M \times M$ identity matrix and $\beta_{k,t} \in \mathbf{R}^M$.

Figure 1: Dynamic Topic Model

Results:

Our final model consists of 15 topics. These topics were discovered from the "Title" field in the registry records which were divided amongst 17 time slices, one for each year based on the "First Received" study variable. Table 2 lists the top words in each topic. The physician author on the research team (AC) analyzed these words to provide human interpretation. We also categorize the topics as "generic" or "topical" based on their disease or organ specific focus or not.

Table 2: Human interpretation based on top words from model

Topic Number	Top words in Topic	**Human Interpretation / Description**	Generic(G) vs. Topical (T)
0	'study', 'trial', 'clinical', 'controlled', 'randomized', 'tive', 'patients', 'pilot', 'evaluation'	**Randomized controlled clinical trials**	G
1	imaging', 'using', 'ultrasound', 'guided', 'study', 'monitoring', 'infants', 'sleep', 'atrial'	**Imaging and monitoring studies**	T
2	'treatment', 'therapy', 'life', 'disorder', 'depression', 'quality', 'disorders', 'stress', 'patients'	**Mental health**	T
3	patients', 'treatment', 'safety', 'efficacy', 'study', 'severe', 'term', 'infection', 'hepatitis'	**Chronic infections and inflammatory disorders**	T
4	'patients', 'acute', 'syndrome', 'coronary', 'artery', 'tissue', 'analysis', 'stroke', 'response'	**Cardiovascular disorders**	T
5	'patients', 'function', 'training', 'device', 'heart', 'failure', 'brain', 'exercise', 'disease'	**Heart and Brain conditions**	T
6	'cancer', 'breast', 'patients', 'women', 'prostate', 'therapy', 'chemotherapy', 'treatment', 'ovarian'	**Breast and prostate Cancer**	T
7	'surgery', 'versus', 'post', 'patients', 'knee', 'cardiac', 'postoperative', 'block', 'following'	**Surgical interventions**	T
8	'cancer', 'patients', 'advanced', 'metastatic', 'cell', 'lung', 'tumors', 'combination', 'carcinoma'	**Advanced cancer treatment**	T
9	disease', 'patients', 'chronic', 'pulmonary', 'refractory', 'relapsed', 'cell', 'lymphoma', 'leukemia'	**Hematologic malignancy, chronic pulmonary disease**	T
10	'care', 'based', 'health', 'patient', 'program', 'management', 'outcomes', 'improve', 'weight'	**Health management**	T
11	‘diabetes', 'intervention', 'activity', 'patients', 'type2', 'children', 'adults', 'control', 'insulin'	**Diabetes**	T
12	'safety', 'study', 'healthy', 'subjects', 'efficacy', 'evaluate', 'tolerability', 'dose', 'pharmacokinetics'	**Vaccines, early stage clinical trials**	T
13	'risk', 'high', 'blood', 'cells', 'patients', 'transplantation', 'stem', 'liver', 'bone'	**Stem cell and organ transplantation**	T
14	pain', 'treatment', 'stimulation', 'patients', 'chronic', 'versus', 'emergency', 'back', 'induced'	**Pain management**	T

Topic Evolution:

Figure 2 describes the topic evolution over 17 years of registered trials. The topics are labeled T0 to T14 with human interpretation. To understand evolution of topics by year, we plotted the mean topic proportion for every time slice from the final model (Figure 2a). Similarly, we plotted the number of documents in the topic for every time slice (Figure 2b). From these plots, we can infer that the number of trials registered in the ClinicalTrials.gov have dramatically increased since its inception in 2000 and in particular since mandatory registration in 2008. (Figure 2b)

As can also be seen from Figure 2a, the document topic proportion varies over time which describes its evolutionary path. This variation would be expected due to changes in research focus and funding priorities and other factors such as finding effective treatments over the course of time.

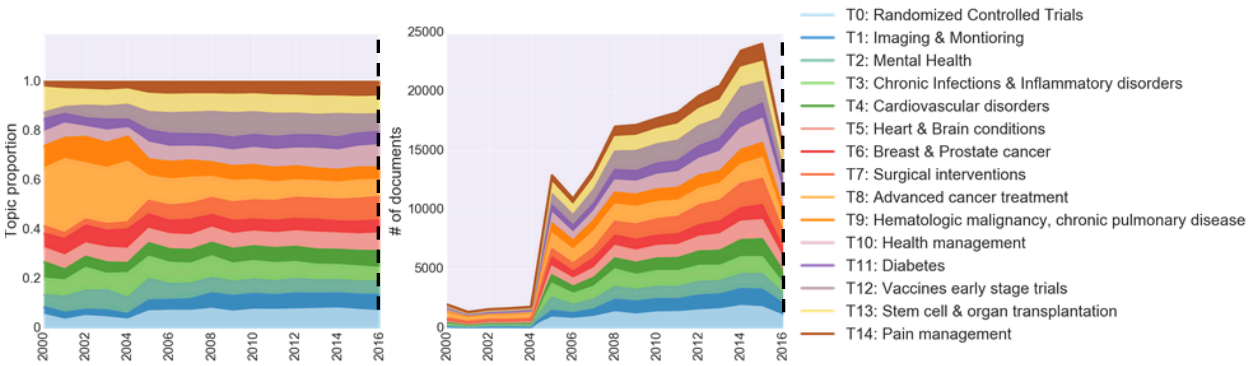

Figure 2: (a) Topic Proportion (b) Documents per topic (- – – – incomplete data for 2016)

For example, hematologic malignancy (T9), and advanced metastatic cancer trials (T8) are being conducted proportionately less in recent years when compared to 2005 or before. (Figure 2a) This may be reflective of a change in funding priorities or due to standardization of treatment options in recent years. Yet at the same time, surgical interventions (T7) and heart and brain conditions (T5) related trials as well as health management (T10) and pain management trials (T14) seem to have gained focus perhaps reflecting a health priority in these areas. [13]

Yet at the same time, the topic of "randomized controlled trials" (T0) is relatively constant, particularly since the beginning of mandatory trial registration requirement in 2008, thus reflecting its "generic" nature.

Insights from metadata in registry records

There are several insights that are derived from matching metadata fields (by document title index) in the registry records with the model output of per-document topic distribution. We describe below our insights from matching each metadata variable in Table 1.

Conditions: First, the very obvious generic theme of Randomized Controlled Trials (T0) is associated with variety of diseases or conditions. Figure 3 describes "Conditions" metadata associated with each topic.

There are also several focused themes in the registry that occur in related diseases or conditions, e.g. psychiatric disorders, heart condition, cancer, women's health, diabetes, postoperative issues, chronic pain etc. For example, topic T2 is predominantly trials on psychiatric conditions. Insights that can be derived from a topic analyses is the degree to which these conditions have been studied in more or less proportion.

In topic T2, Stress and Major Depressive disorder is the focus of a considerable fraction of trials, whereas manic episodes as part of a bipolar disorder seem to be under-represented in this corpus. Of note, schizophrenia is also infrequently studied in trials in this topic, which may reflect the lack of new drug classes to treat this condition.

As part of topic T3 are found chronic infections and inflammatory disorders. The high representation of trials related to the Human Immunodeficiency Virus (HIV) and hepatitis C reflects the major advancements achieved in treating those conditions. New drugs for HIV have turned a lethal disease to a chronic, manageable condition, and now allow cure of the majority of hepatitis C.

Topic T6 has to do mainly with prostate and breast cancer (Figure 3). These are amongst the most prevalent cancer types. Besides its prevalence, the development of biologic treatments for breast cancer may account for its high representation in the trials registry. The long lasting debate over the approach ("to treat or not to treat") for prostate cancer is reflected in its high representation in this topic as well. This topic also includes studies on malaria. Whereas quinine, an anti-malarial drug, has recently been suggested as a treatment for breast cancer [14–16], its representation in this topic is a reminder of a data-driven statistical approach. Being one of the most prevalent diseases affecting mankind, it probably would have justified its own topic, however the volume of research on this topic reflects the under-representation of the developing world population in clinical trials.[17]

Diabetes-related trials dominate topic T11. Of interest, cystic fibrosis, with a prevalence several orders of magnitude lower than diabetes, is intensively studied. This might be owing to the advancement[18] in genetic diagnosis of this condition and its high burden on the healthcare system in terms of hospitalization and antibiotic-resistant infections.

The health management topic (T10) also has a big focus on infant and child health as well as diabetes, mental health and health care costs, all national health and millennium development goals.[19–23] It is interesting to note that most vaccine efficacy and safety trials (T12) are conducted in healthy subjects with a large number of them being influenza and meningococcal vaccine trials. (Figure 3)

Figure 3: Diseases or Conditions in discovered "topics" (size of words reflect the frequency in metadata "conditions" field)

Study Types: The registry consists mostly of "interventional" trials, but there are "observational" and patient registry trials as well. However, diabetes topic area (T11) is an all interventional topic.

Source: This field in the registry defines the source (or the study sponsor) of the trial. Please see Figure 4 for the description that follow. There are few topic areas that are worth noting here. First, a larger proportion of advanced metastatic cancer trials (T8) and hematologic malignancy trials (T9) are sponsored by the National Cancer Institute (NCI). Pharma companies dominate the vaccine efficacy and early stage trials as do they for diabetes (T11), chronic infections (T3). A majority of depression and mental health trials (T2) are conducted at the Veterans Affairs (VA) hospitals and surprisingly none of the pain trials (T14) are conducted in the US. Most health management trails are conducted in geographical locations with high concentrations of academic medical centers – Boston Medical Center,

North Bronx Health Care Network, Ohio State University. Massachusetts General Hospital, Fred Hutchinson Cancer Research Center, Memorial Sloan Kettering Cancer Center, Children's Hospital of Philadelphia top the list for stem cell transplantation trials (T13) and Torrent Pharmaceuticals, Pfizer, Bard and others for conducting sleep monitoring trials (T1).

Figure 4: Source of trials in discovered "topics" (size of words reflect the frequency in metadata field - source)

Gender: Most topic areas contain trials that are conducted in both male and female gender. One exception is vaccine early stage trials (T12) that has a large proportion of male only participants.

Is_FDA_Regulated: Most topic areas have a mix of trials (regulated or not by FDA). However, all trials in health management (T10) topic may be unregulated.

Is_Section_801: All cardiovascular disorder trials (T4) have mandatory reporting requirement according to section 801 of FDAAA and penalties exist according to the current legislation in failing to do so.

Number of study arms: In general, most topics consisted of trials with an average of 2 study arms. However, some diabetes trials (T6) have reported up to 22 arms. Chronic infection trials (T3) and cardiovascular trials (T4) seem to have a higher number of study arms in general (up to 5 study arms).

Overall status: Diabetes (T11), Vaccines (T12) and Imaging and monitoring trials (T1) have the most overall status as "completed". Please see Figure 5 below.

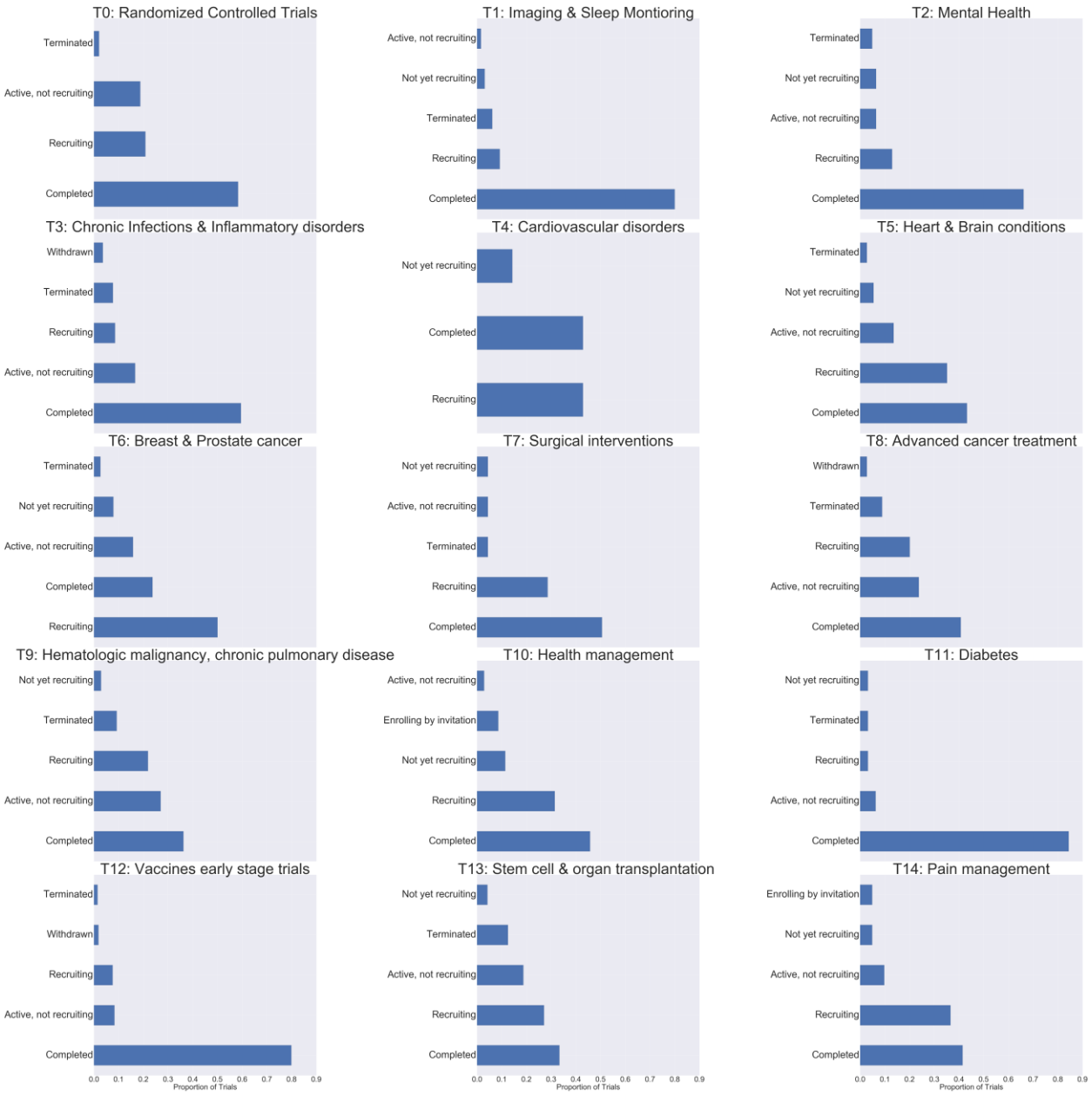

Figure 5: Overall status of trials in each discovered "topic"

Discussion

We used dynamic topic modeling to explore the ClinicalTrials.gov registry records database. Using this method, we discovered the overarching themes in this long running clinical trials registry and results repository. While other studies have evaluated this important resource, their analyses has been more descriptive.[1,24] We believe ours is the first study to conduct a purely data-driven analyses of the ClinicalTrials.gov records. As such the insights gained from these analyses come from the "unknown" because they discover the hidden characteristics of the registry corpus. However, our study has some limitations. First, we selected 15 as the number of thematic areas for our final model

based on initial experimentation. It is possible that with a different number of topics, the model may produce finer or coarser thematic areas and other hidden variables. An alternative approach to address the issue of number of topic selection is a purely data-driven approach such as the hierarchical Dirichlet process. [25] Second, we used "Title" of the study trials for our corpus. It is possible that if another field like NIH agencies existed in the dataset, it could be leveraged to discover similar topics. Third, we set our model hyper-parameter (alpha in Figure 1) to discover "peaky" distribution of topics, i.e. for a study trial to express predominantly a single topic. It is possible that with different values of this parameter, more uniform topic distribution may be discovered. Lastly, despite the clinical expertise, there is some subjectivity to human interpretation of top words in each topic. However, we believe our model is a good start for further improvement. Nonetheless, the findings from this study reflect the evolution of the corpus of ClinicalTrials.gov between 2000 and 2016.

Conclusion

Our study sheds a unique perspective on this important but less utilized health science resource - ClinicalTrials.gov. First, there are trials of generic nature that study health care and health related outcomes. However, they are fewer in number and may not be FDA regulated. Then there are disease or organ focused trials that mainly focus on broad categories but have trends that reflect change in perhaps the research, funding and national health priorities. Of note are advanced and hematologic cancer trials that have downward trended in recent years. Yet, other topics of national health interest such as general health and pain management trials have gained momentum, however they are less likely to be FDA regulated. Interestingly, there are important and prevalent psychiatric disorders that are missing from the mental health trials mix and vaccine trials are largely conducted in healthy subjects with large proportion in males.

References

1. Zarin DA, Tse T, Williams RJ, Carr S. Trial Reporting in ClinicalTrials.gov - The Final Rule. N Engl J Med. 2016 Sep 16;

2. Cepeda MS, Lobanov V, Berlin JA. From ClinicalTrials.gov trial registry to an analysis-ready database of clinical trial results. Clin Trials Lond Engl. 2013 Apr;10(2):347–8.

3. Holman L, Head ML, Lanfear R, Jennions MD. Evidence of Experimental Bias in the Life Sciences: Why We Need Blind Data Recording. PLoS Biol [Internet]. 2015 Jul 8 [cited 2016 Sep 21];13(7). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4496034/

4. Haidich AB. Meta-analysis in medical research. Hippokratia. 2010 Dec;14(Suppl 1):29–37.

5. Vale CL, Rydzewska LHM, Rovers MM, Emberson JR, Gueyffier F, Stewart LA. Uptake of systematic reviews and meta-analyses based on individual participant data in clinical practice guidelines: descriptive study. BMJ. 2015 Mar 6;350:h1088.

6. Gopalakrishnan S, Ganeshkumar P. Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare. J Fam Med Prim Care. 2013;2(1):9–14.

7. About us | Cochrane [Internet]. [cited 2016 Sep 21]. Available from: http://www.cochrane.org/about-us

8. Michelson M. Automating Meta-Analyses of Randomized Clinical Trials: A First Look. In: 2014 AAAI Fall Symposium Series [Internet]. 2014 [cited 2016 Sep 16]. Available from: http://www.aaai.org/ocs/index.php/FSS/FSS14/paper/view/9100

9. Blei DM. Probabilistic Topic Models. Commun ACM. 2012 Apr;55(4):77–84.

10. Blei DM, Lafferty JD. Dynamic Topic Models. In: Proceedings of the 23rd International Conference on Machine Learning [Internet]. New York, NY, USA: ACM; 2006 [cited 2016 Sep 20]. p. 113–120. (ICML '06). Available from: http://doi.acm.org/10.1145/1143844.1143859

11. Tasneem A, Aberle L, Ananth H, Chakraborty S, Chiswell K, McCourt BJ, et al. The database for aggregate analysis of ClinicalTrials.gov (AACT) and subsequent regrouping by clinical specialty. PloS One. 2012;7(3):e33677.

12. Gibbs AL, Su FE. On Choosing and Bounding Probability Metrics. Int Stat Rev. 2002 Dec 1;70(3):419–35.

13. Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research. Mil Med. 2016 May;181(5):397–9.

14. Jiang P-D, Zhao Y-L, Deng X-Q, Mao Y-Q, Shi W, Tang Q-Q, et al. Antitumor and antimetastatic activities of chloroquine diphosphate in a murine model of breast cancer. Biomed Pharmacother Biomed Pharmacothérapie. 2010 Nov;64(9):609–14.

15. Shen Q, Qiu L. Reversal of P-glycoprotein-mediated multidrug resistance by doxorubicin and quinine co-loaded liposomes in tumor cells. J Liposome Res. 2016 Aug 31;1–30.

16. Karthikeyan S, Hoti SL. Development of Fourth Generation ABC Inhibitors from Natural Products: A Novel Approach to Overcome Cancer Multidrug Resistance. Anticancer Agents Med Chem. 2015;15(5):605–15.

17. Mosenifar Z. Population Issues in Clinical Trials. Proc Am Thorac Soc. 2007 May 1;4(2):185–8.

18. Irons JY, Petocz P, Kenny DT, Chang AB. Singing as an adjunct therapy for children and adults with cystic fibrosis. Cochrane Database Syst Rev. 2016 Sep 15;9:CD008036.

19. Khambaty T, Callahan CM, Perkins AJ, Stewart JC. Depression and Anxiety Screens as Simultaneous Predictors of 10-Year Incidence of Diabetes Mellitus in Older Adults in Primary Care. J Am Geriatr Soc. 2016 Sep 19;

20. Smith SL, Shiffman J. Setting the global health agenda: The influence of advocates and ideas on political priority for maternal and newborn survival. Soc Sci Med 1982. 2016 Oct;166:86–93.

21. Tiwari S, Bharadva K, Yadav B, Malik S, Gangal P, Banapurmath CR, et al. Infant and Young Child Feeding Guidelines, 2016. Indian Pediatr. 2016 Aug 8;53(8):703–13.

22. Mathews S, Martin LJ, Coetzee D, Scott C, Brijmohun Y. Child deaths in South Africa: Lessons from the child death review pilot. South Afr Med J Suid-Afr Tydskr Vir Geneeskd. 2016 Sep;106(9):851–2.

23. Furmaniak AC, Menig M, Markes MH. Exercise for women receiving adjuvant therapy for breast cancer. Cochrane Database Syst Rev. 2016 Sep 21;9:CD005001.

24. Ehrhardt S, Appel LJ, Meinert CL. Trends in National Institutes of Health Funding for Clinical Trials Registered in ClinicalTrials.gov. JAMA. 2015 Dec 15;314(23):2566–7.

25. Teh YW, Jordan MI, Beal MJ, Blei DM. Hierarchical Dirichlet Processes. J Am Stat Assoc. 2006 Dec 1;101(476):1566–81.