Developers Spec for Clinicaltrial system

For patients with chronic terminal illness they are often willing to push for their latest treatments which will increase their chances for cure or improvement of quality of life. Often the Doctors/oncologist get to a stage with traditional approved drugs where they are no longer effective. The patient will then have to contemplate trying an experimental drug within a clinical trial. Often the Doctors are unable to assists greatly with this process and the patients have to seek help and resources elsewhere. There are quite a few portals online which have a database of clinical trials all modelled off the main clinicaltrials.gov database which also has an API.

Some of these include http://www.centerwatch.com/clinical-trials/listings/

Trialreach.com etc etc

Understanding the clinicaltrial.gov directory

The link below outlines a snapshot of what some of the records include for each trial

https://clinicaltrials.gov/ct2/show/study/NCT01988896?term=NCT01988896&rank=1#locn

- Basic trial info
- Status (closed, open, recruiting etc)
- Trial Number
- Trial Type (interventional, observational etc)
- Trial Phase (0,1,2, 3,4 or combo)
- Drugs involved
- Condition
- Outcomes
- Enrolment
- Locations
- Start and estimated finish
- Different arms of the trial
- Inclusion and exclusion criteria (is the patient eleigble to enrol in the trial?)

When a person is searching for a trial they will consider a few practical issues regarding the trial which will be based on:

- Location
- Frequency of treatments and follow up

A further analysis will involve looking if the person is eligible for the trial

- This will involve looking at trials which are open and recruiting
- It will also require an analysis of the eligibility and exclusion criteria

Much of the data is quite rigid however you have some text blocks within the inclusion and exclusion criteria as well as drug synonyms and descriptions.

Some engines use NLP to evaluate and extract eligibility and exclusion criteria, they will also run a question survey of users in order to gain their health information which could partially match up to inclusion exclusion criteria – see www.emergingmed.com which help filter out the irrelevant trials.

Solutions

Trial Ranking

A main issue behind the problem of clinical trial searching is that other solutions do not rank the trials meaningfully. In my work with cancer patients I have developed an insight into how to find and isolate the top trials. This is a long process where I learn and find data online.

There are many different types of trials all with different goals and visions. Some trials may be:

- New drugs seeking to significantly improve overall survival and potential cure
- New drugs with the goal to modestly improve survival and or/ patient quality of life (Reduce side effects)
- New drug combination of old drugs to improve potency of drugs
- New delivery and improvement of older drugs – dosing changes or regiment change
- etc

One of the main utilities of this tool will be a ranking of trials via popularity. Most of the top trials are spoken about in various areas online.

- Patient forums
- Blogs
- Scientific literatures
- News articles

The goal would be to build a feature set which will allow for a collection of a signal for each trial where they can be tested via health condition. This can work well with some of the top trials

Other features that can be generated include:

- Sponsorship and funding of the trial
- Academic hype around the drugs involved
- Similarity score to other popular trials
- New drug/combo or not
- Success of drug in other cancers…

Exploring Trials

Another main issue is the display of information about the trial where normally the trials are displayed as search results like google or any other web engine.

The benefit of trials is that they are highly rigid in their theory and mechanism with many trials sharing and overlapping mechanism. Displaying the information more graphically according to these grouping and hierarchy's will help people:

- Get a better understand of the trial landscape
- Find similar trials if the user is ineligible – they could be eligible for other which are similar

- Learn about and explore the various approaches

Another area that is lacking is the content of the trial. Most trials will just strip out all the generic content in clinicaltrials.gov no one is really generating interesting curated content for trials. If they are it is not in a centralized systematic fashion but rather in articles or blogs. It would be cool if we can scrape the best content related to that trail from the web in order to display on our content page for that trial

Similarity Function

Often some of the best trials are either full up and closed for enrolment or for some reasons the patient is not eligible. It would be nice if a patient can view similar trials in this case.

Similarity can be via mode of action of the trial or the drugs/pathways involved.

User Experience

- Patients hits landing page and is asked to enter condition i.e. Colon cancer, breast cancer, diabetes, Alzheimer's etc (in the hack we should just focus on 1 –colorectal cancer)
- They then arrive at the main exploring page to discover and find links to individual clinical trial pages.

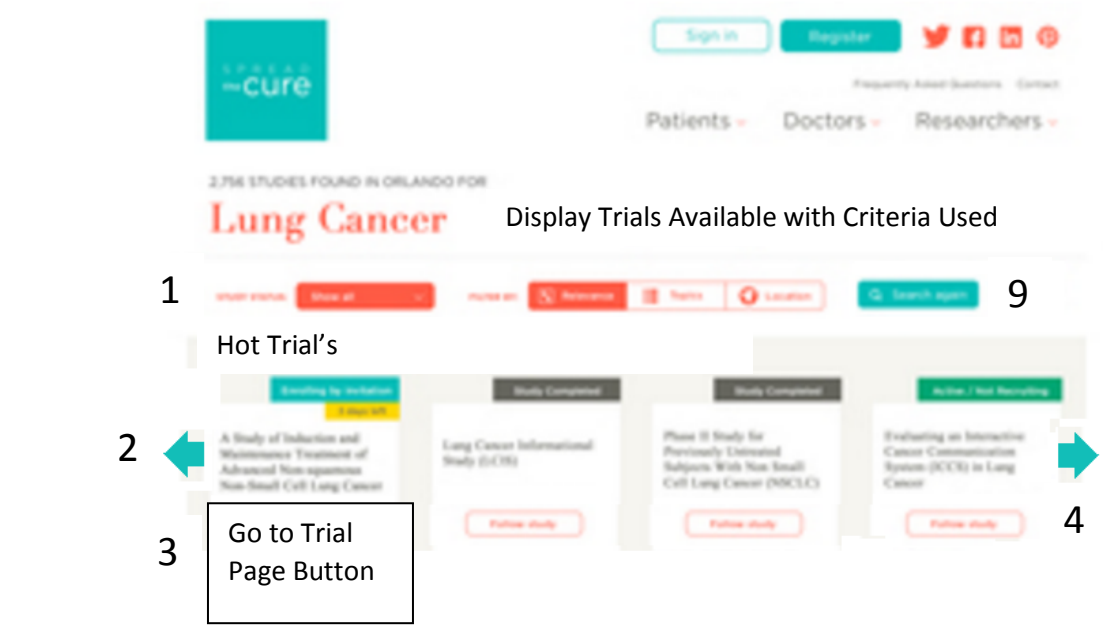
A small horizontal bar will be displayed so patients can filter trials according to:

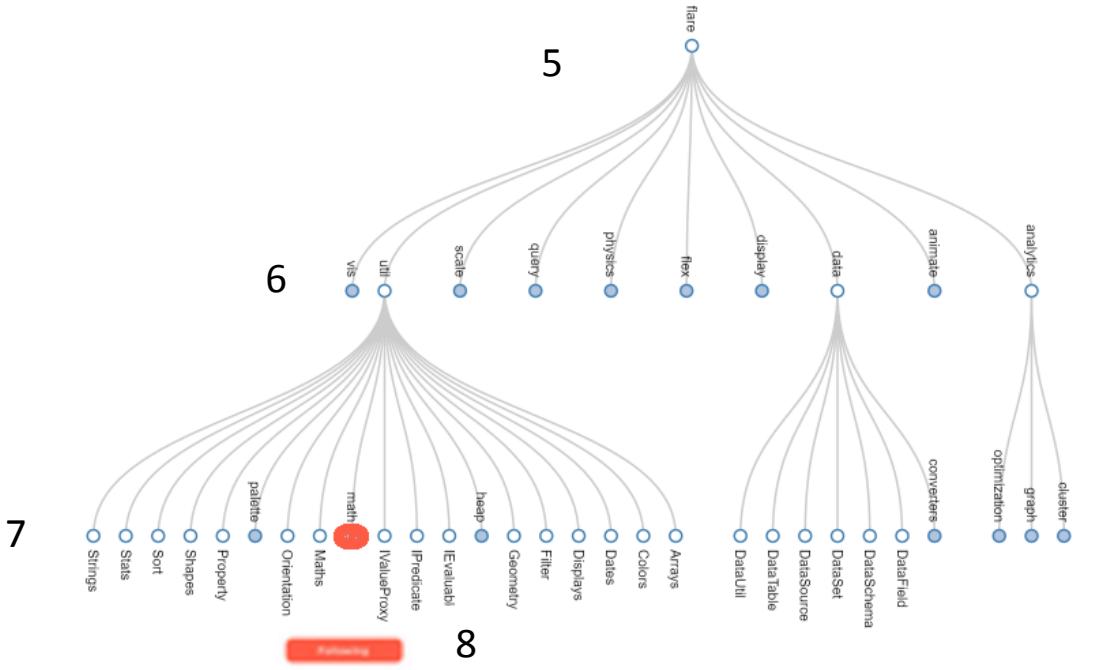
- Phase
- Randomized or not
- Placebo?
- Link for patients to login and enter further health info to filter the eligibility criteria

The top section will consist of a bar displaying the top trials ranked in order where they can be scrolled horizontally. Only small snippets of information will be displayed about those trials

See http://ctdesign.devpost.com/submissions

The drug co Eli Lilly ran a completion 2 years ago asking designers to design a new clinical trials display site. These entries are all have creative common license which we can use.





1. This is the horizontal filter bar with dropdowns for phase
2. This area displays only the top few trial – based on our threshold criteria or a top 20. Initially the top trial is the selected option. Can navigate/scroll via arrow buttons
3. Under the top trial option which is selected there will be two buttons
    a. Go to trial page – which will link to that trial content page
    b. View similar trails – goes to new page which only displays the very similar trials in rank order of similarity via mechanism
4. View trial visualization button – this will cause the tree to move so that trial is displayed. Also the orange circle near no.8 will move to that trials node
5. Trial Visualization tree – this is a responsive tree which groups and classifies according to class or mechanism of the treatment

6. Sub category of treatment – i.e major class (immunotherapy, chemo, targeted, other). This will always be displayed and will not collapse
7. 2nd level sub category (can even be up to around 4 sub categories
8. Button to view trial pages or similar trial page
9. Button to enter in health information to further filter based on location and eligibility

By each node you can also put the numbers of trials contained within that node.

Development Breakdown – For main page

Backend

Trial Classification into Tree group, with main drug mode of action

- Scrape Clinicaltrial.gov data for data about Colorectal Cancer (DB Clinicaltrials.gov)
    - Drugs, with trial number and synonym correction

intervention_type	**intervention_name**	**NCT_id**	**updated_names**
Drug	Atezolizumab	NCT02788279	Atezolizumab
Drug	Cobimetinib	NCT02788279	Cobimetinib
Drug	Regorafenib	NCT02788279	Regorafenib
Drug	Selumetinib and Medi4736	NCT02586987	Selumetinib
	Selumetinib and Medi4736	NCT02586987	Medi4736
Biological	Young TIL	NCT01174121	Young TIL

- Build drug synonym dictionary with all drugs used in trials (DB drugbank http://www.drugbank.ca/)
- Extract and build ATC values for each drug (if there is) (DB drugbank)
- Extract drug target and uniprot id number (if there is) (DB drugbank)
- Manually classify drugs into main and sub class (if there is a sub class)
- Modify Sanger Database (includes targets and its general pathway) to add uniprot ID
- Import the general pathway data into the drug data base

<u>Sample Record for Each Drug</u>

Note not all drugs have full records – if drug has not been approved previously it is likely that it will have no ATC or target info in database. There is currently a limit which I set to 5 targets, this should be enough to cover all the main targets for each drug.

drug	bevacizumab
drugbank_ID	DB00112
ATC_1	ANTINEOPLASTIC AND IMMUNOMODULATING AGENTS
ATC_2	ANTINEOPLASTIC AGENTS
ATC_3	OTHER ANTINEOPLASTIC AGENTS
ATC_4	Monoclonal antibodies
ATC_5	L01XC07
ATC_6	NA
ATC_7	NA

target_1	Vascular endothelial growth factor A
target_2	Low affinity immunoglobulin gamma Fc region receptor III
target_3	Complement C1r subcomponent
target_4	Complement C1q subcomponent subunit A
target_5	Complement C1q subcomponent subunit B
uniprot_ID_1	P15692
uniprot_ID_2	O75015
uniprot_ID_3	P00736
uniprot_ID_4	P02745
uniprot_ID_5	P02746
pathway_1	NA
pathway_2	NA
pathway_3	NA
pathway_4	NA
pathway_5	NA
in_drug_bank	yes
Category	immune system
immue_system_target	transport_immune_cells

Build the Trials Classification Database

- Read in each trail and drug combination
- Import all the Categories, sub categories and Pathway data from the drug DB
- Using pre-determined tree like decision – classify the trial based on the drugs it uses based on current scientific approaches to treatment
- Import general pathways if there are as well from Pathway_1 to Pathway_5 in drug DB

NCT_id	main_cat	sub_cat_1	pathway_1	pathway_2	pathway_3
NCT02788279	immune system	immuno + targeted	PD1/PDL1		
NCT02586987	immune system	immuno + targeted			
NCT01174121	immune system	immuno + chemo	CTLA4		
NCT02350673	targeted	NA	EGFR		

Trial Ranking Function

For now, this is a simple function which queries a search engine API with various search combinations. A simple demo for a few trials was done with Google, which worked well. However, Google blocks multiple heavy searching so only a few trials can be done at once. Alternatively, we can start paying Google/Bing/Yahoo or use another open search API.

A lot of sophistication can be built with this however initially we should keep simple for demo. Crawling specific sites with an NLP engine, combined with searching academic literature using machine learning, could help improve the ranking function.

In additional feature later on will be to use the system to find the best content to import into the content section

Main Data Bases

- Drug data base
- Clinical trial database
    - Mechanistic categories
    - Ranking/classification

Frontend

Hardest challenge will be implementing the d3.js tree. Obviously over time various level of sophistication and presentation can be added for improved UX. A basic demo and functionality will suit for demo.

Cool D3.js interactive data

http://www.nytimes.com/interactive/2012/11/02/us/politics/paths-to-the-white-house.html

http://www.nytimes.com/interactive/2012/02/13/us/politics/2013-budget-proposal-graphic.html?_r=0

http://bl.ocks.org/mbostock/4063550

Other things that will be required will be implementing the display of selected top trial data clinicaltrial.gov data etc etc

Task List for Demo

Before Hack (hopefully)

- Code currently can generate drug data base and classification into its tree category
- Clean up the code and ensure working correctly
- Run on entire Colorectal Cancer Set
- Manually categorize the drugs which are not pre classified
- Categorize colorectal cancer set – approx. 2000-5000 records
- Classify the trials into type of trial – if there is time

Priorities for hack

- Design and build features for ranking function for trials
- implementing d3 tree
- Implement main page design
- Demo of content page

Pluses

- Filtering function
- Similarity Function
- Use only main arm of trial for drug selection per trial

- Improve design of content page
- Add a login and ability for user to add a trial to their basket for review with Dr
- Improve categorization
- Improve ranking
- Add visual features to D3 diagram
- Build additional custom made content pages (currently curated 5 pages), based on ranking order