

[Kayla's comments in purple – Jonny's comments in red]

Hi Jonny,

Here's some feedback (questions, suggestions, food for thought) upon reading your document and spending a little time browsing around the internet on the topic. Some of the questions you might be able to answer easily, some may not have answers. Looking forward to a fruitful project! :)

How does our project differ from existing projects like EmergingMed?

Emergingmed and many other like trial search engines offer the ability to query clinicaltrials.gov as well as add another layer in filtering not offered in clinicaltrials.gov this is the issue of eligibility and exclusion criteria for the trial. The engine will query the patients medical history and use NLP to try and match this to the inclusion/exclusion criteria. Often however if a patient is flexible about trial location they will have many results 100+. The problem and main issue then is trying to filter down from their, which emergingmed does not help.

What, in your experience helping your friend, was the actual process that patients go through in order to find clinical trials? Where are the pain points? (-> How can we make this better.)

The process of further narrowing down the trial will start superficially with inclusion/exclusion criteria, which the emergingmed or other engine filters will miss. Then with the remaining trials 20-50 its about understanding the science behind the trials and analysis of the data available online this will come from:

- Results from earlier phases (ASCO)
- Academic papers
- Company presentations
- news articles
- patient social networks
- research institutes
- etc

Are all (or at least most) clinical trials on clinicaltrials.gov? yes, the vast majority both US and outside are listed there. there are other engines as well which are specific for other areas - like the the EU, UK, Australia etc etc

How many trials are typically considered by a patient? depends on many things like

- logistics and practical situation - are they willing to travel overseas or interstate
- types of treatments they are open to
- Access to info about trials
- willingness and drive of patient to enroll in trial

Why would a patient sign up for a trial with more modest goals?

- lack of knowledge
- logistical constraints
- other trials are closed for recruitment
- Active recruitment and advertising to patient about that trial
- Today around 70% of trials are delivered within the Dr's network - i.e. the Dr will push patient towards trial
- Also Science is uncertain there may be a case where the modest goal trial actual could outperform

Is a trial's "popularity" (measured how exactly?) a meaningful sign of its goodness/relevance? For example, it might simply have good advertising or enroll many patients or be for a more common cancer. As another example, if it's talked about a lot on a forum it could be that there's a recent scandal (Bad). (Sentiment analysis could help us distinguish the latter, if that were desired. It might be tricky on this kind of data involving disease though since a review of a trial can be great but be classified automatically as negative since it contains disease descriptions, like "I had really bad nausea from the side effects but it caused me to go into remission!")

the method proposed in the doc I sent is very primitive, the logic is that popular and hot trials are spoken about the most - which after a bit of testing I found this to be more or less accurate. However as you mentioned I trial can be mentioned in a bad context. Also there may be conflicting data - where initially there is a lot of hype and promise to the trial and then it fails. There are really many different layers one can build to this ranking algorithm which I have thought about - I agree sentiment analysis will be a big part of designing a robust ranking system in this case. as well as

- Mentions of the trial in certain locations
- sentiment analysis in patient forums, academic literature
- ASCO abstract where they release trial results data - and NLP of these summaries

My preference for the short time we have would be to use search engine querying and build features based on results returned, then we can look at the locations of the mention. If we have time we can then potentially do some NLP of the ASCO data release where trial results are published - this is often the first source where a trial becomes very 'hot' it also is semi-structured with a similar format and reporting metrics. i.e. cohort A which contained patients with KRAS mutations vs cohort B with KRAS wild

type had a median OS(overall survival) of 10 months and 16 months respectively and a PFS (progression free survival) of 6 months and 12 months...

see <http://meetinglibrary.asco.org/content/171295-176> for example abstract of a phase 1 trial (each phase will offer different levels of data)

I really like the similarity focus. Should also be able to search by clinical trial and then see similar ones (with the option of adjusting certain search parameters -- an obvious example would be searching for a particular clinical trial, then searching for similar ones that are specifically in the enrollment stage).

Yes, i think this also is a very powerful feature and also a useful tool. For example if the popular trial is full or the patient does not meet the criteria or even does not want to travel they can find similar trials easily.

I would be wary of using too much patient forums / blogs data. They sometimes have good and useful information, but they also often have crazy people making medically inaccurate statements, which if we pass on can sink our project's reliability (and be potentially dangerous). If we do use this kind of data it should be very carefully.

agreed! Hopefully there will be enough data that the outliers will become insignificant

Seems like there should be more focus on patients providing info that is relevant to exclusion/inclusion criteria. For example which chemos (if any) they have had before, etc.

Yes and no. The goal here is not just to find the one or two best trials for the patient. It is also as a learning platform to patient to get more info about trials and discover the different options available. If we provide the filter for eligibility too quickly the patient may miss out on learning about what is out there first. From a UX perspective I feel it is important for them to first enter condition or something simple which then leads them to the main trial results page - after then can filter, which is also super important. In terms of time however for the hack I would prioritize the ranking algorithm vs the filtering of trials - that the innovation here - filtering is not. But I do still hope we will have time to build the eligibility filter.

What do you mean by "Most trials will just strip out all the generic content in clinicaltrials.gov no one is really generating interesting curated content for trials."? The general UX for other trial engines is like this - search, then filter and present list of trials. You click on a trial and you get all the clinicaltrials.gov info about the trial i.e. <https://clinicaltrials.gov/ct2/show/NCT01988896> basic facts about the trial. However there is so much more good content one can deliver about each trial (which is also an important part of the project and design work) Have a look at a wiki page I put together for the same trial I did for some of my patients and compare the difference in the info <http://www.cancertrialwiki.com/clinical-trial-list:nct01988896> - I also made a brief summary video as well for one of my patients.

User experience -- should not have too much focus on creating an account etc. for the hackathon. This is a good idea in general (for saving results, patient privacy issues, etc.) but not data-oriented and relatively standard for non-data developers who can implement this later. Also, when showing judges/presenting, don't want to waste time on creating an account. Can just enter search information (which may include detailed info). **agree**

User experience -- I recommend scrolling vertically instead of horizontally. (Why do you specifically suggest vertically?) - **this is so that there will be enough room for the interaction with graphic below - but we can mock up with the designer and see what works best.**

I assume users (patients / family members / oncologists) will only be interested in one (or a very few) diseases/cancer types, yes? **generalist oncologist and other Dr will potentially be interested in multiple. Also we have not touched on this but we have within cancer there are also many buckets - certain trials will be more beneficial to certain subtypes of cancer, we can discuss this issue too.**

Can we learn from Medivizor?

Medivizor is great. I actually met Tal Givoly the CEO a few times when the company I work for was considering investing in that startup. I know the platform well. There are definitely things we can learn from. However I suspect this would be done further down the line. The main utility is to investigate user interactions with medical content and build user profiles and try and match relevant content to this different profiles. i.e. if we see a person is browsing only trials of type X, we can then customize and send the new trials and content based on this info. etc etc.