|  |  |
| --- | --- |
| Meeting Minutes 10 | |
| Date | 23 December 2019 |
| Start Time | 2:30pm |
| End Time | 4:30pm |

|  |  |  |
| --- | --- | --- |
|  | Agenda | Follow Up Action |
| 1 | Meeting with NCCS  Backend:   * Work together with NCCS to drop unimportant features from dataset * Clarify logic of code   Frontend:   * Gather feedback persona and prototype mock-up * Clarify workflow process of doctors and financial consultants * Clarify whether there are any chart/ visualisation preferences from NCCS |  |

Notes from meeting:

**Backend**

* **Clinical**
* Can drop the types of cancer treatments that are not consistent e.g Extensive\_intraductal etc. or does not affect the number of children (check with Dr. Wong on the hierarchy of the importance of the data)
* Sometimes there is no data for chemotherapy because the patients will go through those treatments in private hospitals. Then suddenly got data for radiotherapy because they come to SGH. But for the purpose of our data model, can drop those.
* Start of chemo - end of chemo. Not important. Can treat it as a yes/no on whether they got attend therapy.
* You don’t become a T1N0 from last diagnosis then “upgrade” to a T1N1 or T2N0 from this diagnosis. If it does it is a recurrence. The diagnosis cancer staging won’t upgrade, will just follow the first diagnosis. Unless the cancer cells appear in the opposite breast then there is other naming for it.
* Try to analyse the exact number of node positive first (3/10, 4/10, 7/10). If not possible/ got time constraint then do N1 (0-3), N2 (4-9), N3(10 nodes or more/ 1 big bulky nodes) etc.
* Minimum refer to the T1,2,3,4 or N1,2,3 (basic grouping) advance grouping is T1N0 stage 1 etc. However, there is no individualism. Highly accurate but also highly imprecise. So he hopes that we don’t do just the minimum
* FISH ratio: NCCS side will clean
* Number of stitches etc. does not matter even though each stitch affects the amount. But they want to find out on e.g this person will need chemo, etc. no matter what, do not need to look at cycles we want to look at completeness for chemotherapy
* Draw dates of surgery etc. for cost analysis throughout time period
* Can drop the dates but let them know cause they might want to correct any mistakes.
* The math for Kaplan Meier (time analysis data) is the same. But they do it based on how they currently divide based on t1, n1. But in our case, we are exploring other ways of dividing such as 1cm, 2cm. So we might have other ways of clustering. 2.5cm and 4.5cm is different, but in their current calculator, they are treated the same. (Difference of the survivability calculator we are doing is number of tumours, sizes, details of cancer and any amount of inputs)
* Treatment date can drop but you need the diagnosis date as day 0. Surgery Date is important - day where cancer is removed e.g last day. **Breast cancer is not cured without surgery.**
* Different people have different start date, and they have different interval (longer/ shorter follow up). Kaplan Meier will stack up all those into the curve.
* RT technique will be a big factor to the cost. Higher risk cancer attracts more advanced radiotherapy, more volume of radiotherapy sessions, etc.
* Last seen is needed (for Kaplan Meier), so cannot drop. Cannot use the death date. It is how the Kaplan Meier curve works. Make no assumption that patients will contribute any information than from when we last saw her 4 years ago.
* Death code: look at disease-free survival and overall survival.
* TNM can be used to form stage 1a, stage 1b, 2a, 2b, etc. Stage 1a can then be grouped into 1,2,3.
* T2N2M0 if u give them neoadjuvant chemo before surgery, everything can melt away??///
* Take the worst case (e.g if she has non-invasive cancer T1 but treated as T2N2?) because if gives us a good % of the inherent risk and it works so cancer melts away
* If patient only have M/N and no T. Just put them as unknown. And clustering algo will already cluster the data into something.
* Logical rules: compare T to T, N to N and M to M.
* Keep breast attracts radio therapy, etc.
* Don’t keep breast need to have reconstruction costs

----------------------------------------------------------------------------------------------------------------

* **Bills**
* service date is 9999: can just drop. If have a reasonable start date, then can just use that one.

----------------------------------------------------------------------------------------------------------------

* **Persona & Workflow**
* Doctor part is fine
* Financial part will be forwarded to financial consultant only to immediate finance and not for the future (5 years) and complication that may arise. Doctor will help ask the finance side to give what they usually cover and then to show what they are expecting.   
    
  Cost side will be gone through by the finance consultant, hoping consultant to go through the cost and up to 5 years etc. as doctors usually wont be involved due to time constraints and lack of knowledge. Doctors may briefly go through the pricing by looking at possible choices at different stages. E.g you dont know when they want remove breast or want reconstructions etc. This could be projected for finance side

Arrange to do biopsy, then brief the biopsy results, go staging etc., nurse do blood taking etc. then Dr. Wong will brief on surgery details, then you will need chemotherapy, hormone therapy, surgery, breast conservation 4 months to 6 months period, within every 3 weeks check blood, after 6 months, check on radiotherapy. Wish for a way where they can predict how long they will undergo treatment and cost.

Any stage above the doctor will go through survivability rate, cost else maybe after given more information doctor can go through. More information cost will be more accurate and lesser. E.g indian 84 year old, 4 nodes positive tumor. Not just use Kaplan Meier but also the Ai data. Ai should filter the data by clusters not just do Kaplan Meier.

Assume surgeon is first time surgery with enough information to tell the patient what to do next for persona and now he is sitting beside him to discuss the next procedures.

The application will be used by many doctors at many stages. E.g. Surgeon use then they will refer patient to chemo specialist → chemotherapy doctor use then refer to radio specialist → radiotherapy doctor use (at any stage the patient can ask about their survival rate and cost so that would be the purpose of our application)

-------------------------------------------------------------------------------------------------------------------------------

* **UI**
* Suggested inputs: TNM, stage, ELPL Her2 (MUST). The rest can play around with it. Can have size also, can have assumptions. Must be some ordering of the things
* (Input page) Interface needs to be a little bit cleaner. He cannot tell which are the selected ones. Not immediately apparent that the “blue” highlighted one is the selected one.
* He think doctor/patient summary is good.
* He wants the like popup box for each data point.
* Filters: cause we already drop a lot of things, there’s not much things to filter. Filter can be grade 1,2,3. Its either we put grade 1,2,3 in inputs or filter. Might not be that necessary to have filters.
* The math behind the curve must be kaplan meier. But for visualization, can use a smoother curve. The colour and confidence interval.
* How to scale to demonstrate the impact of treatment.
* He wants to enter “tick to include radiotherapy, tick to include chemotherapy) → then it will include a scaling of 0.5. → but for this we will need to have a smooth curve.
* E.g. Radiotherapy vs something the curve will reduce cancer by 50%. And this is projected on the curve. If the doctor untick the radiotherapy option, then the curve drops further (survival drops). Or if untick the radiotherapy, then the difference in the curve is not that much, but the cost is like $10000, so the patient can choose whether they want to go or not.
* At every time/data point, can read off and store as x-y.
* The patient-summary, he leave it to us to design the UI. (for the 2-5-10 years one also he leave it to us to design)
* He imagine that when we enter the UI, everything will be ticked. Based on AI results. Then the doctor can slowly untick the filters to look at each individual treatment cost etc.
* Remember to do adjusted value for the cost. Cost-adjustment
* Confidence interval becomes bigger and bigger
* The furthest we can go is to build unsupervised classifier, and we give different groupings rather than fixed TNM, then we run the regressors on these classifiers. → accuracy would be to compare it against the current TNM.
* Segregate data to training and testing data sets. Cross-validation. If we don’t do validation in a robust manner, then the results would not be very good.
* Summary dashboard: is ok
* Predict NHS. They have the R code that powers the predict NHS page. They have smaller dataset of 500
* Read up on predict NHS. he wants minimally that, and better.

<https://breast.predict.nhs.uk/about/technical/publications>

Justin

* Must do the prediction for outcomes first so we can assess the rightness.
* In selection of features, felt that need to narrow down to a specific persona where the patient has minimally TNM, etc and set of features to be able to assess his risk, etc → the doctor inputs those values and the model will churn out the cost →
* What are the features the doctor will key in (e.g. if we can have breast reconstruction then the model will only have data on these.): TNM, Her2, etc. To refine even more can use size-precise and number of nodes.
* E.g. if the patients are 3 years into the treatment vs a patient who just found out about the cancer. → can divert our energy to the beginning part of the treatment. Because by the time 3 years have passed, then most of the costs have already been incurred. So if 3 years patient, doctor will just input all the things, then he will just read off from 3rd year onwards for the graph. He feels that it is not necessary for us to recalculate everything for a 3 year patient and then have a day 0’.
* If we have the number of chemotherapies as an input option, then we need to take into account all the medicine etc.
* Can keep as many variables as we find useful. But Dr Wong have asked lingxing to keep columns that are only useful for predictive model, but those dropped columns may be useful for descriptive model.
* If our model have Y1, Y2, Y3, Y5, Y10, Naturally there will be lesser people who would fill the 10 year population: the graph would just have larger confidence interval at 10 year mark because there is less data to support it.
* Outcomes: either you die or you don’t die. Either you relapse or u don’t relapse