|  |  |
| --- | --- |
| Meeting Minutes 9 | |
| Date | 13 December 2019 |
| Start Time | 1:30pm |
| End Time | 3:30pm |

|  |  |  |
| --- | --- | --- |
|  | Agenda | Follow Up Action |
| 1 | Update progress of backend and frontend |  |
| 2 | Clarify doubts with NCCS |  |
| 3 | Gathering of new requirements |  |

Notes from Meetings

**Team Frontend**

1. Is there any color scheme to follow? Does our design have to match SGH website design?

2. Do we need the singhealth/SGH logo in our UI? (or some rules)

3. Some scenarios /use cases where they use the calculators

4. Any charts that is a must have besides the Kaplan Meier?

5. Any information that patients usually ask about/ wanna know about that would be good to show if we can

* Patients would want something with pictures

Doctors would want more details.

Not everyone will receive certain treatments.

Got a lot of combinations of treatments but just choose the top few to show.

* Can look at the usual/ common combination of chemo - she have to send to us (can look at bill also)

Combinations are quite standard. If want to show the top few choices, then it would be better to show the combinations.

Based on the markers (ER, HER), then they will decide on the treatment

Not very helpful for doctors to know the combinations of treatments because they can just tell from the markers

**Team Backend**

*Issues faced on Data Cleaning*

Bills

1. Possible for us to take only the NCCS code: cannot because sometimes they come in for surgery and chemotherapy, but they overlap with KKH, SGH, etc. as they follow up with different hospitals.
2. Use mapping to map vars??? - e.g. hand surgery
   * E.g. table operation room 1 is a generic code so it can be any surgery.
   * Work in batches, or use sql to udo prefiltering and then joining
3. End date before start date issue
   * Entered by humans so format is different (e.g. 7 may/ july 5)
4. If is blank means no chemotherapy? - yes
5. Map with bill because everything will be inside the bill if patient received anything.
6. Only can combine/ compress??? for patients that come to the hospital for one day (cannot for patients who stay for months)
7. They use gross payable
8. Inflation is usually done at the end for NCCS.
9. Some patients get surgery long after diagnosis date - so shouldn't go too granular - just stick to 1 2 5 10 years
10. Recurrence: cancer cells not there already but come back
11. Alive: disease is always there
12. Survival rate (Kaplan Meier): alive or not alive
13. Other events: e.g. patients have heart problems because of the treatment. Not related to cancer but because of the treatment they have heart problems.
14. Other than survival there are other ways to calculate how effective treatment is → but we shouldn't think about that because it will be more complicated
15. How to reduce the number of data we are looking at:
    * Important remarks column that is free text - text analysis?
    * Bills: Net.exclude.GST can remove
    * Keep: gross, service, service cost, total cost, billed\_qty
    * Doctor… and Service Cost already inside total cost but sometimes they don’t use total cost because not everything is inside the total cost
    * Generally use gross
    * Quantity x [ ] = Gross (has already been multiplied)
16. Service VS Admit Date
17. Can ignore Service Description (unless you wanna map back to see what it is about. Only for understanding)
    * They use the code to define whether needed or not
18. Usually map based on short-text.
19. Why they never send us so many things LOL
20. How to minimise data?
21. Cannot drop service code and only use service short text. Because not everything is in service short text
22. Drop the generic ones
23. Get mapping v2 from SGH

Clinical

* From patient’s perspective in can be different stages
  + If use machine learning to use cancer-related info and non-cancer related info to classify the predictions (different weights) → some of the variables are not inside the data dictionary
    - DOB: start from 2028. Change to 1928. Match with dx age?
  + Match with Age also inaccurate
  + 1889 cannot key in???
  + Use beginning age of treatment or current age? Based on first diagnosis age can calculate back to their age now
  + Hx others
  + Column S: can ignore
  + Many colnames with status behind
  + HisCode\_WWXX (column BA (lx computer)): Additional info other than the T2???? More like what they see in the history
  + First use P\_stage. Only use the c\_stage when they patient want go for adjuvant
  + Tumor size: T1 T2 T3 no size TX
  + Multicollinearity issues
  + Percentage can drop. Usually will just look at negative/ positive to decide treatment
  + % dont really look at it
  + Fish ratio/Fish variable → More for clinician→ is it used frequently? No - but the answer is not certain - need to reconfirm
  + Margin and Margin\_calc: Use the exact size → margin\_mm
  + Cols that are just 100% blanks - information not captured. e.g. If the column for size is blank then they will usually look at the T stage. They will process if in a different way
  + Size\_precise cannot use - use in another way
  + Menopause/Postmenopause more impt
  + HSr\_replacement - 20 is weird num - out of sorts - drop the row?
  + Date\_For\_IDFS: derived information. But can drop
  + TNM collapsed: dont need look at that - summary stage of what you would tell the patient
  + Triple negative: HER, ER, etc all negative
  + Mets\_dx: where has the cancer spread to (cancer spread to other parts of the body)
  + Dose and fractions can drop? → Drop.
    - Not everyone have dose and some patients have alphabets for dose
    - IORT, Boost - 4000 + boost is diff from 4000
    - Because there are different types
  + Can tell what treatment they receive from the bills
  + From billing translate to
  + Bills have a higher resolution of data and more accurate
  + RT - Radiation Therapy