

Minutes from Maja and Catherine's meeting on Monday 21st January 2019, Oxford

Data Management Plan:

We discussed how to progress on the DMP and what should be included it.

- Maja will look for some examples from research project funding applications and forward to Catherine.
- DMP should include clear rules for data input e.g. how and when missing values are used.
- DMP should clearly and comprehensively describe the migration process and any issues that arose during it and any decisions that had to be made during transfer, along with relevant interpretations and definitions
- naming conventions should also be described here. for files, for variables etc.
- Catherine to look into rules around data sharing between partners, CHASE and other third parties etc to add to plan.

Data

- in the past some of the data has been estimated. e.g. when the split of first time users and repeat users wasn't known it was sometimes estimated. This needs to be flagged in the dataset! e.g. with a separate variable indicator for the source of these variables: from the partners or estimates.
- Additionally there is the question of what does 'first use' of a contraceptive method even mean? This should be defined somewhere clearly.
- in the past some of the data was for multiple day clinics. This means the row by row data is not comparable. But this is actually true anyway, the data can only be reasonably compared by month/quarter or year. But we could add a variable to indicate whether the clinic was single or multiple day.
- Adding the county for each clinic location for better location analysis. This is easy going forward but could be tricky for the data already collected.

Data collection/input

It would be good to have a clearer understanding of the local partner's process of data input. What are they doing now for Chase Africa as well as what they are doing for others e.g. MOH. This is to get a better idea of how much effort they are already putting into this for CHASE's sake, whether they find this useful for their own record keeping.

It would also be good to know where the potential dangers are on their side e.g. do they collect paper records over weeks and then input them all at once? Would it be possible for them to input the records immediately after a clinic?

Catherine to check with Henry. Update: "Partners record patient individual data in paper books at the clinics and these are then logged as a data summary on Excel in the office at the end of the clinic. The MOH receives the raw paper records which they then use to compile their own reports. Partners do not share the data summaries they produce for CHASE with the MOH. Partners produce the data summary shared with CHASE specifically for CHASE and CHASE only. This means it should be relatively straightforward to switch to a different report format provided we can train them to use Google Forms" additionally "there is no internet access on clinic so not an option. Data would be submitted at the end of a clinic from the office as before."

So from the partners point of view they could have three options for passing on their information:

- status quo - they add rows to the existing excel table and send it over to CP once every quarter, who migrates the data manually

- a new table - i.e. a tidy table with no subtotals and no calculations in the spreadsheet. this would be better data management practice, would not significantly change the amount of effort required on their part, and would make the data migration safer and more streamlined.
- Google forms - depending on what their data input process is this could be more, same or less effort than what they are doing now. Regardless, the transition would represent a one off investment of time, but it is difficult to estimate how much. It would be great if we were able to pilot it!

We started on the Google form and Catherine will complete it.

Data storage

We came up with a proposal for the data storage that allows for all data collection options described above. This is to have a single Google sheet with separate worksheets for

- the migrated data - this sheet is protected, it was input once and should not be changed
- the responses from the Google forms - this sheet is also protected and data is only automatically input from the forms
- new inputs - if data continues to be provided via an excel table (current or new) this is where Catherine inputs it into a separate sheet, which is not protected from input.
- main consolidated data - this spreadsheet takes data from all three sheets and merges it into one table, additionally this is where any calculations are performed, not before. Again this is a spreadsheet that should be protected, as once it is set up there is no need for further intervention.

Dashboard

Instead of the previous suggestion to use shiny, we now think that Google Data Studio provides a better solution to represent the data and analysis in the form of a “dashboard”. The main reason is that after we set it up Catherine will be able to administrate it without Maja’s help, and will be able to create new dashboards herself as well. Google Data Studio provides a great platform for setting up a dashboard which would allow CHASE to share findings easily internally and with partners with automatic reports that update as new data is input.

- overview of important stats
- quick visualisations of all variables, original and derived
- easy access to subsets of data table, easy to export
- modularity: separate dashboards can be set up for each individual partner, and a main more comprehensive dashboard to be used by chase Africa.

Next steps

1. Data collection/input: Catherine will set up the Google forms.
2. Maja will set up the Google sheets.
3. Both will consider what sort of information should be included in the dashboards.
4. Both will start preparing sections of the DMP.