Feedback Notes

1. The subject_id in both ODK and MIMIC data should have the same format. The ODK uses '_', meanwhile the MIMIC uses an integer data type.
2. Since there have been 3rd and 4th doses (boosters) of COVID-19 vaccination, the COVID-19 vaccination data in ODK should display 5 categories: '0', '1 dose', '2 dose', '3 dose', and '4 dose'.
3. If the purpose of ODK is to find the analysis related to vaccination, there should be variables related to the type of vaccines of each jab, for example: mRNA, viral vector, or protein subunit.
4. Since the COVID-19 virus has been mutated several times, the ODK should have a variable related to the variants or subtypes of the virus from patient sequencing.
5. Data pseudonymization tends to create limitations and biases in data analysis. For several patients, the admit_age from ODK does not match to calculation of age from deathtime and dob from MIMIC data. There should be a consideration related to important data for analysis (e.g., age).
6. The ODK data contains sensitive data, for example, age and location. In the future, should be a strict protocol and more security layers to access this data.
7. Several variables might be redundant in MIMIC, for example, dod_hosp and deathtime.
8. There are several patients with earlier dischtime than admittime. I assumed it was mistakes in filling the data. In the future, there should be a method to prevent mistakes in filling out the form.
9. The clear dictionaries of both databases ODK and MIMIC. Several internet resources have different perceptions related to MIMIC variables. For example, the definition of los in minutes by SUNY Oswego and los in hours by University of Pittsburgh.
10. The sex and gender in ODK data can be ambiguous for analysis, the data should be only contains one of those variables.
11. Several sex in ODK do not match with gender in MIMIC data. There should be clear which one is reliable for analysis in the future and related to the treatment of patients.
12. The ICD 9 diagnosis of patients should be clear about the time or location of the diagnosis.
13. There should be a clear instruction and guidance related to the report. The instruction can make different results for different persons/data managers.
14. In the future, the pipeline of the executive report should be like the scheme below: