# Task# 1 & 2a Deliverable – Arti

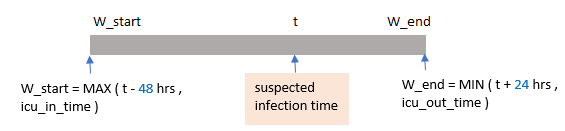
1. SEPSIS ONSET CRITERIA

Output file => sepsis\_onset\_time.csv

Criteria is based on Desautels’s paper (SOFA>=2), implemented in Big Query.

*Taking the initial time of the earliest culture draw or antibiotic administration as the time of suspicion of infection, we define a window of up to 48 hours before this time (limited by time of data availability) and 24 hours after this time (limited by time of departure from the ICU). The SOFA score at the beginning of this window is compared with its hourly value throughout this window; if this*

*hourly value is ≥ 2 points higher than the value at the start of the window, we define the first such hour as the onset of sepsis.*



1. CASE PATIENT SELECTION AND FEATURES :

Output file => Case\_vitals.csv

1. Select icu\_stays where

* sofa\_score change >=2 between W\_start & W\_end window
* and age > 15 years.
* and sepsis\_onset\_time – icu\_in\_time > 18 hours

1. Roll up vitals in 1-hour bins for patients identified in step-1. Measurement value was averaged within 1-hour bin.
2. Add age & gender info.

@Dhruv : Please consider filtering out patients from Case\_vitals.csv that have less than x observations. Including these patients can impact model quality as we don’t have meaningful data for these patients.

Eg : there are 560 patients (out of 1346) that have >= 18 sample ( where 1 sample is 1-hour data) .

1. PATIENTS DETECTED AS CASE PATIENTS BY OTHER SEPSIS CRITERIA

Output file => Sepsis\_superset\_patientIDs.csv

This file contains list of patients that have been identified by at least of the four criteria – Angus , Martin , CDC\_Sepsis , Explicit\_ICD\_code

@Satish – any patient in Sepsis\_superset\_patientIDs.csv or sepsis\_onset\_time.csv should ideally not be part of control patient set.