

Appendix 4. Documentation for data clean up for ICU

The following steps were used to ensure those who received ventilator use are counted for in the ICU use, and to calculate the length of stay for individuals requiring ICU care.

The variables in **bold** were created, whereas other CAPITALIZED variables are from the original data source.

1. Create new variable **ICUSTART**: if ICU2 = 1, take earliest date from: ICUSTART, ICUEND, VENTSTART2. If there are no dates then show a blank.
2. Create new variable **ICUEND**: if ICU2 = 1, take latest date from: ICUSTART, ICUEND, VENTEND. If there are no dates then show a blank.
3. Create new variable **ICUCHECK**: This variable will show 1 if ICU2 = 1, ICUSTART – ICUEND = 0, and count (ICUSTARTDATE, ICUENDDATE) = 1, else 0. Same rationale as VENTCHECK in A2.
4. Create a new variable **ICUEND2**: if HOSP=1, ICUCHECK=1, and count(HOSPADMITDATE, HOSPDISCHARGEDATE) > 1, take max of HOSPADMITDATE/HOSPDISCHARGEDATE, otherwise keep ICUEND. The purpose of this variable is to check if there is a ICUCHECK flag, then see if there is an HOSP end date from original dataset. If yes, then assume that for ICUEND2, if not then take the current ICUEND.
5. Create a new variable **ICUSTART2**: if HOSP=1, ICUCHECK=1, AND COUNT(HOSPADMITDATE, HOSPDISCHARGEDATE) > 1, take min of HOSP*DATE, otherwise ICUSTART. The purpose of this variable is to check if there is a ICUCHECK flag, then see if there is a HOSP start date from original dataset. If yes, then assume that for ICUSTART2, if not then take the current ICUSTART.
6. Create a new variable **ICU_LOS**: If ICUEND2-ICUSTART2 <= 0 AND ICUCHECK = 1, then “X”, else ICUEND2-ICUSTART2. IFERROR then “X”. This variable is used for the LOS for patients in the ICU. In the first part of that logic statement, it says if there is a flag, and the dates are not in the correct order, then LOS is not calculated and these individuals are excluded.