**SQL QUERY**   
-- Extract rows with SOFA >= 2 for sepsis detection

WITH sofa AS (

    SELECT stay\_id,

           starttime,

           endtime,

           ROUND(sofa\_24hours, 2) AS sofa\_score,

           ROUND(respiration\_24hours, 2) AS respiration,

           ROUND(coagulation\_24hours, 2) AS coagulation,

           ROUND(liver\_24hours, 2) AS liver,

           ROUND(cardiovascular\_24hours, 2) AS cardiovascular,

           ROUND(cns\_24hours, 2) AS cns,

           ROUND(renal\_24hours, 2) AS renal

    FROM physionet-data.mimiciv\_derived.sofa

    WHERE sofa\_24hours >= 2

),

-- Extract suspicion of infection data

suspicion AS (

    SELECT subject\_id,

           stay\_id,

           antibiotic\_time,

           culture\_time,

           suspected\_infection\_time,

           suspected\_infection

    FROM physionet-data.mimiciv\_derived.suspicion\_of\_infection

),

-- Apply exclusion criteria to filter patients

filtered\_admissions AS (

    -- Select first admission only

    WITH first\_admissions AS (

        SELECT \*

        FROM physionet-data.mimiciv\_hosp.admissions

        QUALIFY ROW\_NUMBER() OVER(PARTITION BY subject\_id ORDER BY admittime) = 1

    ),

    -- Exclude pregnancies

    no\_pregnancy AS (

        SELECT fa.\*

        FROM first\_admissions fa

        LEFT JOIN physionet-data.mimiciv\_hosp.diagnoses\_icd diag ON fa.hadm\_id = diag.hadm\_id

        LEFT JOIN physionet-data.mimiciv\_hosp.d\_icd\_diagnoses icd ON diag.icd\_code = icd.icd\_code

        WHERE icd.long\_title NOT LIKE '%pregnancy%' OR icd.icd\_code IS NULL

    ),

    -- Exclude ICU stays > 100 days

    short\_icu\_stays AS (

        SELECT np.\*

        FROM no\_pregnancy np

        JOIN physionet-data.mimiciv\_icu.icustays icu ON np.hadm\_id = icu.hadm\_id

        WHERE TIMESTAMP\_DIFF(icu.outtime, icu.intime, DAY) <= 100

    ),

    -- Exclude ICU stays < 24 hours

    valid\_icu\_stays AS (

        SELECT sis.\*

        FROM short\_icu\_stays sis

        JOIN physionet-data.mimiciv\_icu.icustays icu ON sis.hadm\_id = icu.hadm\_id

        WHERE TIMESTAMP\_DIFF(icu.outtime, icu.intime, HOUR) >= 24

    )

    SELECT \*

    FROM valid\_icu\_stays

),

-- Combine filtered patients with suspicion of infection data

valid\_suspicion AS (

    SELECT s.\*

    FROM suspicion s

    JOIN filtered\_admissions fa ON s.subject\_id = fa.subject\_id

),

-- Combine SOFA and suspicion of infection data to identify sepsis

sepsis AS (

    SELECT s.subject\_id,

           s.stay\_id,

           s.antibiotic\_time,

           s.culture\_time,

           s.suspected\_infection\_time,

           s.suspected\_infection,

           f.sofa\_score,

           f.respiration,

           f.coagulation,

           f.liver,

           f.cardiovascular,

           f.cns,

           f.renal,

           CASE

               WHEN f.sofa\_score >= 2 AND f.sofa\_score < 4 THEN 'Sepsis'

               WHEN f.sofa\_score >= 4 THEN 'Septic Shock'

               ELSE 'No Diagnosis'

           END AS sepsis\_type,

           ROW\_NUMBER() OVER (

               PARTITION BY s.stay\_id

               ORDER BY s.suspected\_infection\_time, s.antibiotic\_time, s.culture\_time, f.endtime

           ) AS rn

    FROM valid\_suspicion s

    INNER JOIN sofa f

    ON s.stay\_id = f.stay\_id

    AND f.endtime BETWEEN DATETIME\_SUB(s.suspected\_infection\_time, INTERVAL 48 HOUR)

                      AND DATETIME\_ADD(s.suspected\_infection\_time, INTERVAL 24 HOUR)

    WHERE s.suspected\_infection = 1

),

-- Extract SIRS patients based on ICD codes

sirs AS (

    SELECT icu.subject\_id,

           icu.stay\_id,

           diag\_icd.icd\_code,

           CASE

               WHEN diag\_icd.icd\_code IN ('99591', '99592') THEN 'SIRS'

               ELSE 'No Diagnosis'

           END AS diagnosis

    FROM physionet-data.mimiciv\_icu.icustays icu

    LEFT JOIN physionet-data.mimiciv\_hosp.diagnoses\_icd diag\_icd

    ON icu.hadm\_id = diag\_icd.hadm\_id

    WHERE diag\_icd.icd\_code IN ('99591', '99592')

),

-- Extract additional fields within 24 hours after ICU admission including new prognostic data

additional\_fields AS (

    SELECT icu.subject\_id,

           icu.stay\_id,

           p.gender,

           p.anchor\_age AS age,

           ROUND(AVG(CASE WHEN ce.itemid = 224639 THEN ce.valuenum END), 2) AS Weight\_kg,

           ROUND(AVG(CASE WHEN ce.itemid = 223900 THEN ce.valuenum END), 2) AS GCS,

           ROUND(AVG(CASE WHEN ce.itemid = 220045 THEN ce.valuenum END), 2) AS HR,

           ROUND(AVG(CASE WHEN ce.itemid = 220179 THEN ce.valuenum END), 2) AS SysBP,

           ROUND(AVG(CASE WHEN ce.itemid = 220181 THEN ce.valuenum END), 2) AS MeanBP,

           ROUND(AVG(CASE WHEN ce.itemid = 220180 THEN ce.valuenum END), 2) AS DiaBP,

           ROUND(AVG(CASE WHEN ce.itemid = 220210 THEN ce.valuenum END), 2) AS RR,

           ROUND(AVG(CASE WHEN ce.itemid = 220277 THEN ce.valuenum END), 2) AS SpO2,

           ROUND(AVG(CASE WHEN ce.itemid = 223761 THEN ce.valuenum END), 2) AS Temp\_C,

           ROUND(AVG(CASE WHEN ce.itemid = 223835 THEN ce.valuenum END), 2) AS FiO2\_1,

           ROUND(AVG(CASE WHEN ce.itemid = 227442 THEN ce.valuenum END), 2) AS Potassium,

           ROUND(AVG(CASE WHEN ce.itemid = 227465 THEN ce.valuenum END), 2) AS Sodium,

           ROUND(AVG(CASE WHEN ce.itemid = 227486 THEN ce.valuenum END), 2) AS Chloride,

           ROUND(AVG(CASE WHEN ce.itemid = 220621 THEN ce.valuenum END), 2) AS Glucose,

           ROUND(AVG(CASE WHEN ce.itemid = 227493 THEN ce.valuenum END), 2) AS BUN,

           ROUND(AVG(CASE WHEN ce.itemid = 220615 THEN ce.valuenum END), 2) AS Creatinine,

           ROUND(AVG(CASE WHEN ce.itemid = 220635 THEN ce.valuenum END), 2) AS Magnesium,

           ROUND(AVG(CASE WHEN ce.itemid = 220645 THEN ce.valuenum END), 2) AS Calcium,

           ROUND(AVG(CASE WHEN ce.itemid = 220602 THEN ce.valuenum END), 2) AS Ionised\_Ca,

           ROUND(AVG(CASE WHEN ce.itemid = 220645 THEN ce.valuenum END), 2) AS CO2\_mEqL,

           ROUND(AVG(CASE WHEN ce.itemid = 220587 THEN ce.valuenum END), 2) AS SGOT,

           ROUND(AVG(CASE WHEN ce.itemid = 220586 THEN ce.valuenum END), 2) AS SGPT,

           ROUND(AVG(CASE WHEN ce.itemid = 225690 THEN ce.valuenum END), 2) AS Total\_bili,

           ROUND(AVG(CASE WHEN ce.itemid = 220659 THEN ce.valuenum END), 2) AS Albumin,

           ROUND(AVG(CASE WHEN ce.itemid = 220228 THEN ce.valuenum END), 2) AS Hb,

           ROUND(AVG(CASE WHEN ce.itemid = 51277 THEN ce.valuenum END), 2) AS MPV, -- Mean Platelet Volume

           ROUND(AVG(CASE WHEN ce.itemid = 50813 THEN ce.valuenum END), 2) AS CRP, -- C-Reactive Protein

           ROUND(AVG(CASE WHEN ce.itemid = 51265 THEN ce.valuenum END), 2) AS PLTC, -- Platelet Count

           ROUND(AVG(CASE WHEN ce.itemid = 51256 THEN ce.valuenum END), 2) AS WBCC, -- White Blood Cell Count

           ROUND(AVG(CASE WHEN ce.itemid = 51250 THEN ce.valuenum END), 2) AS NeuC, -- Neutrophil Count

           ROUND(AVG(CASE WHEN ce.itemid = 51244 THEN ce.valuenum END), 2) AS LymC, -- Lymphocyte Count

           SAFE\_DIVIDE(AVG(CASE WHEN ce.itemid = 51250 THEN ce.valuenum END), AVG(CASE WHEN ce.itemid = 51244 THEN ce.valuenum END)) AS NLCR, -- Neutrophil/Lymphocyte Ratio

           ROUND(AVG(CASE WHEN ce.itemid = 227466 THEN ce.valuenum END), 2) AS PTT, -- Partial Thromboplastin Time

           ROUND(AVG(CASE WHEN ce.itemid = 227467 THEN ce.valuenum END), 2) AS PT, -- Prothrombin Time

           ROUND(AVG(CASE WHEN ce.itemid = 220561 THEN ce.valuenum END), 2) AS INR, -- International Normalized Ratio

           ROUND(AVG(CASE WHEN ce.itemid = 220734 THEN ce.valuenum END), 2) AS Arterial\_pH,

           ROUND(AVG(CASE WHEN ce.itemid = 220735 THEN ce.valuenum END), 2) AS paO2,

           ROUND(AVG(CASE WHEN ce.itemid = 220739 THEN ce.valuenum END), 2) AS paCO2,

           ROUND(AVG(CASE WHEN ce.itemid = 220745 THEN ce.valuenum END), 2) AS Arterial\_BE,

           ROUND(AVG(CASE WHEN ce.itemid = 220750 THEN ce.valuenum END), 2) AS Arterial\_lactate,

           ROUND(AVG(CASE WHEN ce.itemid = 220751 THEN ce.valuenum END), 2) AS HCO3 -- Bicarbonate

    FROM physionet-data.mimiciv\_icu.icustays icu

    LEFT JOIN physionet-data.mimiciv\_hosp.patients p ON icu.subject\_id = p.subject\_id

    LEFT JOIN physionet-data.mimiciv\_icu.chartevents ce ON icu.stay\_id = ce.stay\_id

    WHERE TIMESTAMP\_DIFF(ce.charttime, icu.intime, HOUR) <= 24

    GROUP BY icu.subject\_id, icu.stay\_id, p.gender, p.anchor\_age

)

-- Combine sepsis, SIRS, and additional fields data

SELECT

    s.subject\_id,

    s.stay\_id,

    s.antibiotic\_time,

    s.culture\_time,

    s.suspected\_infection\_time,

    s.sofa\_score,

    s.respiration,

    s.coagulation,

    s.liver,

    s.cardiovascular,

    s.cns,

    s.renal,

    s.sepsis\_type AS diagnosis,

    af.gender,

    af.age,

    af.Weight\_kg,

    af.GCS,

    af.HR,

    af.SysBP,

    af.MeanBP,

    af.DiaBP,

    af.RR,

    af.SpO2,

    af.Temp\_C,

    af.FiO2\_1,

    af.Potassium,

    af.Sodium,

    af.Chloride,

    af.Glucose,

    af.BUN,

    af.Creatinine,

    af.Magnesium,

    af.Calcium,

    af.Ionised\_Ca,

    af.CO2\_mEqL,

    af.SGOT,

    af.SGPT,

    af.Total\_bili,

    af.Albumin,

    af.Hb,

    af.MPV,

    af.CRP,

    af.PLTC,

    af.WBCC,

    af.NeuC,

    af.LymC,

    af.NLCR,

    af.PTT,

    af.PT,

    af.INR,

    af.Arterial\_pH,

    af.paO2,

    af.paCO2,

    af.Arterial\_BE,

    af.Arterial\_lactate,

    af.HCO3

FROM sepsis s

LEFT JOIN additional\_fields af ON s.subject\_id = af.subject\_id AND s.stay\_id = af.stay\_id

WHERE s.rn = 1

UNION ALL

SELECT

    si.subject\_id,

    si.stay\_id,

    NULL AS antibiotic\_time,

    NULL AS culture\_time,

    NULL AS suspected\_infection\_time,

    NULL AS sofa\_score,

    NULL AS respiration,

    NULL AS coagulation,

    NULL AS liver,

    NULL AS cardiovascular,

    NULL AS cns,

    NULL AS renal,

    si.diagnosis,

    af.gender,

    af.age,

    af.Weight\_kg,

    af.GCS,

    af.HR,

    af.SysBP,

    af.MeanBP,

    af.DiaBP,

    af.RR,

    af.SpO2,

    af.Temp\_C,

    af.FiO2\_1,

    af.Potassium,

    af.Sodium,

    af.Chloride,

    af.Glucose,

    af.BUN,

    af.Creatinine,

    af.Magnesium,

    af.Calcium,

    af.Ionised\_Ca,

    af.CO2\_mEqL,

    af.SGOT,

    af.SGPT,

    af.Total\_bili,

    af.Albumin,

    af.Hb,

    af.MPV,

    af.CRP,

    af.PLTC,

    af.WBCC,

    af.NeuC,

    af.LymC,

    af.NLCR,

    af.PTT,

    af.PT,

    af.INR,

    af.Arterial\_pH,

    af.paO2,

    af.paCO2,

    af.Arterial\_BE,

    af.Arterial\_lactate,

    af.HCO3

FROM sirs si

LEFT JOIN additional\_fields af ON si.subject\_id = af.subject\_id AND si.stay\_id = af.stay\_id

ORDER BY subject\_id, stay\_id, suspected\_infection\_time;

1. **Extract rows with SOFA ≥ 2 for sepsis detection (WITH sofa AS (...))**:
   * This section selects patients with a sofa\_24hours score of 2 or higher, an indicator of organ dysfunction associated with sepsis.
   * The columns selected (sofa\_score, respiration, coagulation, liver, cardiovascular, cns, renal) provide details about specific organ systems over a 24-hour period, rounded to two decimal places.
2. **Extract suspicion of infection data (WITH suspicion AS (...))**:
   * Retrieves infection-related information (like antibiotic\_time, culture\_time, and suspected\_infection\_time) for each patient, which is crucial for diagnosing sepsis. This data indicates suspected infections based on recorded medical events.
3. **Apply exclusion criteria to filter patients (WITH filtered\_admissions AS (...))**:
   * Filters patients based on specific conditions:
     + **First admissions only**: Selects only the first admission for each patient.
     + **Excludes pregnancies**: Uses ICD codes to filter out patients with pregnancy-related diagnoses.
     + **ICU stay restrictions**: Includes patients with ICU stays between 24 hours and 100 days to ensure relevance and consistency.
4. **Combine filtered patients with infection suspicion data (WITH valid\_suspicion AS (...))**:
   * Joins suspicion with filtered\_admissions to ensure that only qualified patient admissions are considered for sepsis analysis.
5. **Identify sepsis based on SOFA and infection data (WITH sepsis AS (...))**:
   * Combines sofa data with valid\_suspicion data to identify and categorize sepsis based on sofa\_score:
     + Sepsis (SOFA score between 2 and 4).
     + Septic Shock (SOFA score of 4 or higher).
   * This section orders records by relevant timestamps (suspected\_infection\_time, antibiotic\_time, culture\_time, endtime) and assigns a row number (rn) to prioritize the most accurate sepsis diagnosis for each stay.
6. **Extract SIRS patients based on ICD codes (WITH sirs AS (...))**:
   * Selects patients diagnosed with SIRS (Systemic Inflammatory Response Syndrome) using ICD codes 99591 and 99592. These codes categorize SIRS separately from sepsis, as SIRS can occur without infection.
7. **Extract additional fields within 24 hours after ICU admission (WITH additional\_fields AS (...))**:
   * Gathers detailed vital signs and laboratory results within the first 24 hours of ICU admission. These metrics are critical for assessing the patient’s condition and identifying early signs of sepsis or SIRS.
   * Uses ROUND(AVG(...)) to calculate the average values for each parameter, such as heart rate (HR), blood pressure (SysBP, MeanBP, DiaBP), temperature (Temp\_C), and other biochemical markers.
8. **Combine sepsis, SIRS, and additional fields data (SELECT ... UNION ALL SELECT ...)**:
   * Combines records for patients identified as having sepsis (based on SOFA score and infection criteria) with those diagnosed with SIRS.
   * Uses a LEFT JOIN with additional\_fields to incorporate demographic and clinical measurements.

**Key Points:**

* **UNION ALL**: Combines sepsis and SIRS datasets, preserving all records, and orders them by subject\_id, stay\_id, and suspected\_infection\_time.
* **CASE statements**: Provide classification for sepsis and non-sepsis diagnoses.
* **Aggregation and rounding**: Calculated fields are averaged and rounded to two decimal places to maintain precision in measurements.