**Introduction**

This report outlines the data collection processes and methods employed for gathering sepsis readmission data used for analysis and reporting. The dataset provides crucial insights into patient demographics, sepsis admission and discharge timelines, and subsequent readmissions. The data is sourced primarily from Epic’s Workbench reporting tool, with additional data retrieval options through Snowflake using SQL queries.

**Data Sources and Tools**

1. **Epic Workbench Reporting Tool**
   * **Source**: The primary tool used to extract the sepsis readmission dataset is Epic’s Workbench Reporting Tool. Epic is an electronic health record (EHR) system widely used in hospitals for managing patient data.
   * **Process**:
     + Clinicians and data analysts query the Epic Workbench tool to retrieve patient data for sepsis-related admissions and readmissions.
     + Workbench allows users to generate customized reports based on the required data elements (e.g., patient demographics, sepsis admission, and discharge dates, etc.).
     + Data fields are selected according to pre-defined criteria, and reports are generated for further analysis.
2. **Snowflake Data Warehouse**
   * **Source**: Snowflake is another tool used to retrieve the dataset, with the added benefit of SQL querying capabilities for more complex or custom data extraction needs.
   * **Process**:
     + The dataset can be queried using SQL scripts, which allow for specific filtration, aggregation, and sorting of the data.
     + Snowflake enables data retrieval from large datasets or from databases that are not easily accessible directly from Epic.

**Data Collection Methods**

The following methods are used to collect and extract the sepsis readmission dataset:

1. **Data Querying from Epic Workbench**:
   * Reports are generated using the Workbench tool by selecting relevant data fields. The columns used in the dataset are based on standardized data definitions in the Epic system.
   * The query focuses on sepsis-related patient data, capturing all relevant identifiers (CSN, MRN) and demographic information (age, sex, ethnicity, etc.).
   * Admission and discharge dates related to sepsis are recorded, allowing for the tracking of hospital length of stay (LOS) and subsequent readmissions.
   * Key columns include:
     + **CSN** and **MRN** for unique patient identification
     + **Pt Name**, **DOB**, **Pt Age on Adm**, and **Sex** for demographic details
     + **Sepsis Adm Dt** and **Sepsis Disch Dt** for sepsis-specific admission and discharge dates
     + **LOS** (Length of Stay), **Readmission Dt**, and **Days from Sepsis Disch Dt** to track readmissions and hospital stays
     + **Sepsis Disch Dept** and **Readmission Dx Combo** to understand discharge departments and reasons for readmissions
2. **Data Retrieval via Snowflake SQL Queries**:
   * SQL queries are designed to select, filter, and aggregate data from large datasets.
   * SQL allows for joining multiple tables to capture all necessary data points related to sepsis admissions, discharges, and readmissions.
   * Specific filters can be applied based on variables such as admission dates, diagnoses, payor types, race, ethnicity, and more.
   * For example, a SQL query might filter for patients who had sepsis as a primary diagnosis or those who were readmitted within 30 days of discharge.

**Key Variables Collected**

The data collected for sepsis readmission analysis includes the following key variables:

1. **CSN (Client Service Number)**: Unique identifier for each patient encounter.
2. **MRN (Medical Record Number)**: Unique identifier for each patient.
3. **Pt Name (Patient Name)**: Full name of the patient (typically anonymized or masked for privacy in reports).
4. **DOB (Date of Birth)**: Patient's date of birth, used for age calculations.
5. **Pt Age on Admission**: Age of the patient at the time of admission for sepsis.
6. **Sex**: Gender of the patient (e.g., male, female).
7. **Payor**: The insurance or financial payer type covering the patient's care (e.g., private insurance, Medicare).
8. **Race**: Patient's racial background (used for demographic analysis).
9. **Ethnicity**: Patient's ethnic background (used for demographic analysis).
10. **Sepsis Adm Dt (Sepsis Admission Date)**: The date the patient was admitted with sepsis.
11. **Sepsis Disch Dt (Sepsis Discharge Date)**: The date the patient was discharged after treatment for sepsis.
12. **LOS (Length of Stay)**: The number of days the patient stayed in the hospital during the sepsis admission.
13. **Readmission Dt (Readmission Date)**: The date the patient was readmitted to the hospital after sepsis discharge.
14. **Days from Sepsis Disch Dt**: The number of days between the sepsis discharge date and the subsequent readmission date.
15. **Sepsis Disch Dept (Sepsis Discharge Department)**: The department from which the patient was discharged after sepsis treatment (e.g., ICU, general medicine).
16. **Readmission Dx Combo**: The combination of diagnoses associated with the readmission (e.g., sepsis recurrence, other complications).

**Data Accuracy and Integrity**

* **Epic Workbench Reporting Tool**: Since the data is sourced directly from Epic’s EHR system, the accuracy of the data relies on the accuracy of the inputted clinical information and the integrity of the reporting tool.
  + Validation checks in the Epic system help to ensure that data entered by clinicians is accurate.
  + Manual oversight and regular audits ensure that the correct data is being extracted from the system.
* **Snowflake SQL Queries**: Data integrity in Snowflake relies on the correctness of SQL queries written by data analysts. Careful design of queries ensures that data is pulled accurately from the Snowflake data warehouse and reflects the correct timeframes, diagnoses, and patient details.

**Limitations of Data Collection Process**

1. **Data Completeness**: Some patients may not have complete records in the EHR system, leading to missing or incomplete data for certain variables.
2. **Data Quality**: There may be discrepancies in how some data points are recorded, particularly around subjective fields like race, ethnicity, and diagnoses. Ensuring consistency in these fields is essential for accurate reporting.
3. **Timeliness**: Data extraction depends on the availability of up-to-date information in Epic and Snowflake. Delays in updating records may affect the accuracy of reports.

**Conclusion**

The sepsis readmission dataset is crucial for understanding the patterns of sepsis-related admissions and readmissions. The data is collected primarily from Epic's Workbench reporting tool and Snowflake through SQL queries, providing comprehensive insights into patient demographics, sepsis treatment timelines, and readmission patterns. While the data collection process is robust, there are always challenges related to completeness, consistency, and timeliness that need to be monitored to ensure high-quality reporting.

This dataset is instrumental for clinical and administrative teams to identify potential areas for improving patient outcomes, reducing readmission rates, and enhancing the quality of care for sepsis patients.