**Comprehensive Report on data collection processes and methods.**

**Introduction**

This report outlines the data collection methods and processes used for the Increase Pneumococcal Vaccine Compliance Rate PI Project. The dataset includes patient demographics, admission details, length of stay, discharge information, and vaccine-related data, all of which are crucial for tracking vaccine compliance and identifying trends to improve pneumococcal vaccine administration rates.

The primary sources of data for this project are Epic's Electronic Health Record (EHR) system, with data extracted via the Epic Workbench reporting tool, and Snowflake, a cloud-based data warehouse, using SQL queries to handle more complex data extraction needs.

**Data Sources and Tools**

1. **Epic Workbench Reporting Tool**
   * **Source**: The primary tool for data extraction is Epic's Workbench reporting tool. Epic is widely used by healthcare organizations to manage patient data.
   * **Process**:
     + Clinicians and analysts use the Workbench tool to generate reports based on pre-defined criteria.
     + The focus is on variables that align with the goal of improving pneumococcal vaccine compliance.
     + Data fields selected include patient demographics, vaccination status, discharge information, length of stay, and provider details.
2. **Snowflake Data Warehouse**
   * **Source**: Snowflake is a cloud-based data warehouse that facilitates large-scale data extraction through SQL queries.
   * **Process**:
     + Analysts create custom SQL queries to extract and manipulate data, ensuring that relevant patient and vaccination information is included.
     + Snowflake allows joining multiple tables, applying filters, and aggregating data based on specific variables.

**Data Collection Methods**

1. **Data Querying via Epic Workbench**
   * Reports are generated by querying the Epic Workbench tool, focusing on pneumococcal vaccine-related data.
   * Key data fields selected to track compliance include patient demographics, vaccine status, length of stay, discharge disposition, and attending provider details.
   * The dataset includes:
     + Patient name, MRN (Medical Record Number), and age for identification and demographic information.
     + Admit date, discharge month, and length of stay to track hospital stay duration.
     + Discharge disposition, discharge time, and discharge unit for capturing discharge details.
     + Primary payer, DRG (Diagnosis-Related Group), service, and attending provider for healthcare and billing information.
     + Vaccine compliance data, including vaccine status and vaccine administered during the encounter.
     + Interpreter required, language, and ethnicity for assessing patient needs and ensuring effective vaccine communication.
     + Problem list, HM topics due, and vaccine compliance status to evaluate whether the patient received the vaccine during their stay.
2. **Data Retrieval via Snowflake SQL Queries**
   * Custom SQL queries are used to extract required data directly from the Snowflake data warehouse.
   * SQL queries allow analysts to filter, aggregate, and join data from different tables, ensuring only relevant records are included.
   * SQL queries are designed to monitor vaccine compliance based on various factors, including age, discharge status, and conditions like comorbidities, race, or ethnicity.

**Key Variables Collected**

The dataset includes several critical variables, particularly related to patient demographics, healthcare details, and pneumococcal vaccine compliance. Key variables include:

1. Name: Full name of the patient (anonymized in reports).
2. MRN: Medical Record Number, a unique identifier for the patient.
3. Date of Birth: Used to calculate the patient's age.
4. HAR (Hospital Account Record): Unique identifier for patient encounters.
5. Admit Date: The date of patient admission.
6. Eff Date: Effective date of care, typically the admission date.
7. Discharge Month: Month of patient discharge.
8. Discharge Disposition: How the patient was discharged (e.g., home, transfer, expired).
9. Pt Class: Category of the patient (e.g., inpatient, outpatient).
10. CSN: Client Service Number, another unique identifier for encounters.
11. Age: The patient's age at admission.
12. Race and Ethnicity: Demographic data for inclusion in compliance tracking.
13. Language: Spoken language, important for vaccine communication.
14. Interpreter Required: Indicates if an interpreter was needed for the patient.
15. Discharge Time and Eff Time: Times related to discharge and care dates.
16. Length of Stay: Duration of the patient’s stay in the hospital.
17. Attending Provider: The primary provider responsible for care.
18. Admit and Principal Diagnosis: Primary reasons for hospitalization.
19. DRG: Diagnosis-Related Group, used for billing and reporting.
20. Primary Payer: The primary insurance provider.
21. Vaccine Compliance Status: Indicates if the patient received the pneumococcal vaccine.
22. Vaccine Given on B6 During Encounter: Indicates whether the vaccine was administered during the encounter.
23. HM Topics Due: Health management topics relevant to the care plan.
24. Vaccine Compliant?: A flag indicating whether the patient complied with the vaccine protocol.

**Data Accuracy and Integrity**

* **Epic Workbench Reporting Tool**: Data accuracy is ensured through built-in validation processes, where clinicians are prompted to accurately enter clinical details such as vaccination status and diagnosis codes.
  + Regular audits help maintain data quality, reducing discrepancies in critical fields like vaccine compliance.
  + Reporting through the Workbench tool ensures data is extracted according to the most recent records in the Epic system.
* **Snowflake SQL Queries**: Accuracy is maintained by ensuring that SQL queries are properly designed to extract the correct data. Analysts verify the consistency and completeness of query results before using them for analysis.
  + Parameters are adjusted in queries to account for missing data or outliers, ensuring a robust dataset.

**Limitations of Data Collection Process**

1. **Data Completeness**: Some records may lack key vaccination data or important variables like discharge disposition or race, which could affect the accuracy of compliance calculations.
2. **Data Entry Errors**: Inaccurate or inconsistent data entry (e.g., incorrect vaccine status or diagnosis) may lead to errors in analysis. Although validation protocols help minimize these issues, they cannot eliminate them entirely.
3. **Timeliness**: Delays in updating patient records or generating reports can impact the availability of up-to-date data for analysis.
4. **Patient Privacy**: Patient confidentiality is maintained through anonymization of personally identifiable information (PII) to comply with privacy regulations such as HIPAA.

**Conclusion**

The Increase Pneumococcal Vaccine Compliance Rate PI Project relies on data integration from Epic’s EHR system and Snowflake’s data warehouse. By querying both systems, healthcare providers and analysts can monitor vaccine compliance, identify care gaps, and track patient outcomes. The data collection methods ensure that important variables related to vaccination, patient demographics, and hospital care are included, contributing to the success of the project.

While there are challenges related to data completeness and accuracy, the methodologies used ensure that the final dataset is reliable and useful for enhancing pneumococcal vaccine compliance rates.