### **Script Documentation: Structured Echocardiography Data Extraction**

#### **Overview**

This SQL script extracts structured data from echocardiography reports in the MIMIC-III database, leveraging regex patterns to capture important clinical details. The extracted data includes timestamps, patient vitals, and echocardiography-specific information such as the type of test and technical quality. The script processes free-text echocardiography notes, converting them into structured fields that can be joined with other tables, allowing for more efficient analysis.

#### **Key References**

* MIMIC-III Clinical Database: Echo reports contain both structured and unstructured data. This script extracts specific fields from the free-text portion using regular expressions.

#### **Logic Summary**

* **Text Parsing**: Uses regular expressions (regex) to capture specific fields such as the indication for the test, blood pressure (BP), heart rate (HR), and other relevant clinical data.
* **Charttime Recreation**: The script extracts the time of the echocardiography from the text, as charttime is missing for this category of notes in MIMIC-III.
* **De-identification Handling**: Some numeric values are obscured by de-identified text in MIMIC-III, and the script removes these to return valid clinical data.

#### **Process Steps**

1. **Drop Table**: The script checks if the echo\_data table exists and drops it if present to ensure that the new table creation is successful.
2. **Create Table**: A new table echo\_data is created, which includes:
   * **Patient Identifiers**: subject\_id, hadm\_id, and ROW\_ID.
   * **Date/Time Handling**: The charttime is imputed by combining the chartdate with the time extracted from the echocardiography text.
   * **Indication Extraction**: Extracts the reason for the echocardiography from the notes.
   * **Vitals and Clinical Data**: Extracts height, weight, body surface area (BSA), blood pressure (systolic and diastolic), heart rate, and other structured clinical information.
3. **Regular Expressions**:
   * The script uses regex to capture values from free-text echocardiography reports (e.g., BP, HR).
   * Non-greedy matching (?) ensures that only the desired portions of the text are extracted.

#### **Output**

The script generates a new table, echo\_data, containing the following fields:

* **Patient Identifiers**: ROW\_ID, subject\_id, hadm\_id, chartdate.
* **Recreated Charttime**: Imputed using the extracted timestamp from the text.
* **Clinical Data**:
  + Height, Weight, BSA, BP, BPSys, BPDias, HR.
  + Echo-specific fields such as Indication, Test, Doppler, Contrast, TechnicalQuality.

#### **Example Query**

To retrieve echocardiography reports where systolic blood pressure is over 140 mmHg:

SELECT subject\_id, hadm\_id, BP, BPSys

FROM echo\_data

WHERE BPSys > 140;

#### **Important Notes**

* **ROW\_ID Versions**: ROW\_ID differs across versions of MIMIC-III, so ensure that the correct version is being referenced.
* **Text De-identification**: Some numeric fields may contain de-identified placeholders (e.g., [\*\* Numeric Identifier \*\*]). The script handles these cases by removing the placeholders and attempting to cast the remaining text to numeric values.

#### **Conclusion**

This script extracts key clinical information from unstructured echocardiography reports in MIMIC-III, converting it into structured data for further analysis. The resulting table can be easily joined with other patient-level data using subject\_idor hadm\_id, allowing for comprehensive research and analysis of echocardiography results.