**TB Patient Dataset (Synthetic): Bangladesh**

**Description of the Dataset:**

The dataset contains **20,000 records** of synthetic data representing **Tuberculosis (TB) patients** in **Bangladesh**, including various clinical, demographic, and social variables. The data is structured as follows:

**1. Patient Information:**

* **Patient ID**: A unique identifier for each patient (integer values from 1 to 20,000).
* **Age**: Numeric value representing the patient’s age (range: 0–100 years).
* **Gender**: The gender of the patient, with values **Male**, **Female**, and **Other**.
* **Region**: Geographic region of the patient (Urban or Rural).
* **Occupation**: Occupation type (e.g., **Farmer**, **Laborer**, **Student**, **Worker**, **Other**).

**2. Diagnosis Information:**

* **Date of Diagnosis**: Number of days since the diagnosis was made (numeric).
* **Symptoms**: Symptoms reported by the patient, including **Cough**, **Fever**, **Weight loss**, **Night Sweats**.
* **Sputum Smear Test**: The result of the sputum smear test (either **Positive** or **Negative**).
* **GeneXpert Test**: The result of the GeneXpert MTB/RIF test (either **Positive** or **Negative**).
* **Chest X-ray Results**: Result of the chest X-ray (either **Normal** or **Abnormal**).

**3. Treatment Information:**

* **Treatment Start Date**: Number of days since treatment started (numeric).
* **Treatment Type**: Type of treatment received, either **DOTS** (Directly Observed Treatment, Short-course) or **Drug-resistant TB treatment**.
* **Duration of Treatment**: The length of the treatment in months (range: 6–24 months).
* **Drug Resistance**: Indicates whether the patient has **Drug-resistant TB** (**Yes** or **No**).
* **Treatment Outcome**: Outcome of the treatment (e.g., **Cured**, **Failed**, **Lost to Follow-up**, **Death**).
* **Relapse**: Whether the patient experienced a relapse after treatment (**Yes** or **No**).
* **Mortality**: Whether the patient died as a result of TB (**Yes** or **No**).

**4. Co-morbidities:**

* **Malnutrition**: Whether the patient suffers from **malnutrition** (**Yes** or **No**).
* **Diabetes**: Whether the patient has **Diabetes** (**Yes** or **No**).
* **HIV**: Whether the patient has **HIV** (**Yes** or **No**).
* **Chronic Lung Disease**: Whether the patient has **Chronic Lung Disease** (**Yes** or **No**).

**5. Social and Environmental Factors:**

* **Smoking Status**: Smoking status of the patient (e.g., **Non-smoker**, **Smoker**, **Ex-smoker**).
* **Alcohol Consumption**: Whether the patient is a **Non-drinker**, **Drinker**, or **Ex-drinker**.
* **Living Conditions**: The socioeconomic living conditions of the patient (e.g., **Poor**, **Average**, **Rich**).
* **Access to Healthcare**: Whether the patient has access to healthcare services (**Yes** or **No**).

**6. Geographic Information:**

* **City**: The city where the patient resides (e.g., **Dhaka**, **Chittagong**, **Khulna**, **Rajshahi**, **Sylhet**, and others).
* **Region Code**: Numeric code representing different regions of Bangladesh (values ranging from 1 to 64, corresponding to the various administrative divisions).

**Guideline with Exact References:**

This dataset is based on the **WHO** and **National Tuberculosis Control Programme (NTP) of Bangladesh** guidelines, which provide the clinical framework for managing **Tuberculosis (TB)**.

1. **WHO Global Tuberculosis Report 2022**:
   * **Key Guidelines**: WHO’s guidelines provide information on the epidemiology, diagnosis, treatment, and prevention of TB. The report includes data on TB incidence, mortality, and drug-resistant TB, which helps in shaping the treatment and diagnostic categories in this dataset.
   * **Reference**: World Health Organization (WHO), **Global Tuberculosis Report 2022**. Available at: <https://www.who.int/teams/global-tuberculosis-programme/tb-reports>
2. **NTP Bangladesh Annual Report**:
   * **Key Guidelines**: The **National Tuberculosis Control Programme** of Bangladesh provides national standards for TB treatment, diagnosis, and epidemiological reporting. The guidelines include protocols for **Directly Observed Treatment Short-course (DOTS)** and **drug-resistant TB treatment**.
   * **Reference**: National Tuberculosis Control Programme (NTP) Bangladesh, **Annual Report 2022**. Available at: <https://www.ntp.gov.bd/wp-content/uploads/2024/01/Annual-Report-2022.pdf>
3. **CDC TB Guidelines**:
   * **Key Guidelines**: The **Centers for Disease Control and Prevention (CDC)** guidelines provide comprehensive information on TB diagnosis and management, including the **GeneXpert MTB/RIF** test and **sputum smear microscopy**.
   * **Reference**: Centers for Disease Control and Prevention (CDC), **Clinical Guidance for Tuberculosis Treatment**. Available at: <https://www.cdc.gov/tb/hcp/clinical-guidance/index.html>

**References and Citations for the Dataset:**

Here is the **correct citation** for the sources used:

1. **World Health Organization (WHO)**, **Global Tuberculosis Report 2022**, [WHO Global Tuberculosis Report](https://www.who.int/teams/global-tuberculosis-programme/tb-reports).
2. **National Tuberculosis Control Programme (NTP) Bangladesh**, **Annual Report 2022**, [NTP Bangladesh Annual Report 2022](https://www.ntp.gov.bd/wp-content/uploads/2024/01/Annual-Report-2022.pdf).
3. **Centers for Disease Control and Prevention (CDC)**, **Clinical Guidance for Tuberculosis Treatment**, [CDC Tuberculosis Clinical Guidelines](https://www.cdc.gov/tb/hcp/clinical-guidance/index.html).

**In Summary:**

This **synthetic TB dataset** closely follows the key guidelines and recommendations from the **WHO**, **NTP Bangladesh**, and **CDC**. It represents a broad spectrum of variables that are typically used to track and analyze **Tuberculosis** in Bangladesh, including **clinical outcomes**, **diagnostic tests**, **treatment regimens**, and **comorbidities**. The references provided are based on **official guidelines** that help to ensure the data is **consistent** with current **global and national standards** for TB management.