

Smoking & Cancer Risk Analysis

This dataset comprises detailed records of patients, focusing on the relationship between smoking behaviour and cancer risk. It integrates demographic, lifestyle, environmental, and medical information to support analytical research in Oncology, Epidemiology, and preventive healthcare.

The dataset has been normalized into three relational tables:

- Patients: General demographic and lifestyle attributes
- Habits: Smoking history and exposure levels
- Results: Cancer type and Cancer stage.

1. Patients

The Patients table captures general demographic and lifestyle information for each individual in the dataset. Each entry corresponds to a unique patient and includes key health and lifestyle factors that may contribute to cancer risk. This foundational data allows us to analyse how variables like age, BMI, and diet influence outcomes when cross-referenced with medical data.

Key fields include:

- Patient_ID: Unique identifier for each patient.
- Age: Age of the patient at the time of record.
- Gender: Patient's gender (Male, Female, Other).
- BMI: Body Mass Index — indicator of body fat based on height and weight.
- Physical_Activity_Level: Patient's general physical activity (Low, Moderate, High).
- Diet_Quality: Quality of the patient's diet (Poor, Average, Good).

2. Habits

The Habits table focuses on each patient's smoking history and related environmental factors. By isolating smoking habits and exposure duration, this table supports investigation into how smoking contributes to cancer development. It links directly to the Patients table through the Patient_ID.

Key fields include:

- Patient_ID: Identifier linking to the patients.
- Smoking_Status: Whether the patient is a current, former, or non-smoker(Never).
- Cigarettes_Per_Day: Average number of cigarettes smoked daily
- Years_Smoking: Total years the patient has been smoking

3. Results

The Results table details each patient's medical outcome, including cancer type, stage. This dataset provides the basis for analysing correlations between lifestyle habits and cancer .It plays a key role in outcome-based queries.

Key fields include:

- Patient_ID: Foreign key referencing the patients.
- Cancer_Type: Type of cancer diagnosed, if any (e.g., Lung, Mouth, Throat, None, Other).
- Cancer_Stage: Stage of cancer (I, II, III, IV) or missing if not diagnosed.