# Inpatient Head and Neck Cancer Medication Data

This case-study includes a partial subset of data for head and neck cancer patients. This dataset mainly focuses on the inpatient medication and cancer status of 595 patients.

## Driving Biomedical Challenges

1. Determine the relation between the patient SEER stage of cancer and other covariates, e.g., treatment medications, dosage, frequency, etc.
2. Use text-mining and/or natural language processing methods to extract structured data from the unstructured free-text, e.g., MEDICATION\_DESC, MEDICATION\_SUMMARY.
3. Apply machine learning techniques to automatically segment the patients into clusters and determine how well these derived groupings correspond to patient phenotypes, e.g., cancer stage (SEER).

## Variables

**PID:** codedpatient ID

**ENC\_ID:** coded encounter ID

**Seer\_stage:** SEER cancer stage (0 =In situ, 1=Localized, 2=Regional by direct extension, 3=Regional to lymph nodes, 4=Regional (both codes 2 and 3), 5=Regional, NOS, 7= Distant metastases/systemic disease, 8=Not applicable, 9=Unstaged, unknown, or unspecified). See: <http://seer.cancer.gov/tools/ssm>.

**Medication\_desc:** description of the chemical composition of the medication

**Medication\_summary:** brief description about medication brand and usage

**Dose:** the dosage in the medication summary

**Unit:** the unit for dosage in the **Medication\_summary**

**Frequency:** the frequency of use in the **Medication\_summary**

**Total\_dose\_count:** total dosage count according to the **Medication\_summary**