Technical Test: Quantitative Sciences

Your datasets consist of:

Medication orders

patient_id: patient identifier for internal purposes
external_patient_id: patient identifier for external (client) purposes
order_id: unique identifier of each medication ordered by a health provider
order_date: date in which medication was ordered by a health provider

Medication administrations

patient_id: patient identifier for internal purposes

external_patient_id: patient identifier for external (client) purposes

order_id: unique identifier of each medication ordered for a patient by a health provider

administered date: date in which ordered medication was administered

drug_name: drug administered to patient

Patient demographics

patient_id: patient identifier for internal purposes

gender: gender of patient

age: age of patient at the time of diagnosisrace: racial group classification of patient

Patients

patient id: patient identifier for internal purposes

external_patient_id: patient identifier for external purposes
external_practice_id: practice identifier for external purposes

internal_practice_id: practice (site at which patient receives oncology care) identifier for Internal purposes

diagnosis_date: date on which patient was initially diagnosed

advanced_diagnosis_date: date on which patient was diagnosed as advanced

Practices

practice id: client identifier for internal purposes

external_practice_id: client identifier for external (client) purposes

emr system: software product used by the client

(OncoE = OncoEMR; OncoA = OncoAnalytics; OncoB = OncoBilling)

practice_type: classification of practice into either community or academic bucket

active: status of client

QUESTIONS

- 1. When presented with a new dataset or database, what steps do you generally take to evaluate it prior to working with it?
- 2. Based on the information provided above and the attached dataset, what 3 questions would you like to understand prior to conducting any analysis of the data?
- 3. How would you prep the dataset provided for analysis? Please list steps taken and provide code used to prep the tables for analysis.

Using the datasets you have prepped, please answer the following questions and provide all code/output used. Be sure to include and explain any assumptions made when answering each question.

- 1. What is the average time elapsed between a patient's initial diagnosis date and a patient's first treatment? Does this time vary by gender?
- 2. How many patients are on nivolumab from 2012-2016?
- 3. Using the following risk stratification rules, please summarize the number of high, medium, and low risk patients.

```
HighRisk = Female; any-age; NON_WHITE OR Male; >= 70; NON_WHITE

MediumRisk = Female; >=75, WHITE OR Male; <70; NON_WHITE

LowRisk = Female; < 75, WHITE OR Male; any-age; WHITE
```

4. Please create a visualization that could be used to help a medical researcher? understand how drug prevalence has changed over time.