REQUEST OVERVIEW

Project Name	Help Everyone Actually Really Truly (HEART)
Objective	Model Risk of Heart Failure after Drug Exposure
Target Completion Date	9/5/2024 ☐ flexible ☒ firm
Requester	Dutch Redlake
Summary	This a retrospective cohort study modeling risk of heart failure diagnosis following exposure to one of two drugs, while controlling for various patient characteristics.

DATA SPECIFICATIONS

Population

Include any patient who received either drug A or drug B.

Exclude any data for "test patients" generated as a test of the electronic medical record software.

Input Datasets

There are three input datasets provided in CSV format. Some data cleaning may be required, so the analyst should be on guard for data which deviates from the specifications.

1. Patients
   This dataset contains relevant patient characteristics. One row per patient.

Variable	Type	Notes
study_id	char(6)	Primary key
first_name	char(11)	
middle_name	char(11)	
last_name	char(11)	
suffix	char(3)	
height	integer	Height in centimeters
member	boolean	A value of 1 indicates the patient is a member of the health plan
birth_date	date	

2. Dispenses
   Contains data on dispenses of drugs A and B. One row per patient and dispense.

Variable	Type	Notes
study_id	char(6)	Primary key
dispense_date	date	Primary key
drug_id	char(6)	Identifies drug A or drug B
days_supply	integer	The number of days the amount dispensed will last if taken as prescribed

3. Diagnoses
   Contains data on heart failure (HF) diagnoses (dx) for patients in the cohort. One row per patient per dx.

Variable	Type	Notes
study_id	char(6)	Primary key
dx_date	date	Date of diagnosis
dx_code	char(7)	ICD-10 CM diagnosis code
dx_name	char(200)	Diagnosis code description

Diagnosis codes

The following ICD-10 CM diagnosis codes were included in the diagnoses dataset:

ICD-10 CM Code	Description
I50.0	Congestive heart failure
I50.1	Left ventricular failure
I50.2	Systolic (congestive) heart failure
I50.20	Unspecified systolic (congestive) heart failure
I50.21	Acute systolic (congestive) heart failure
I50.22	Chronic systolic (congestive) heart failure
I50.23	Acute on chronic systolic (congestive) heart failure
I50.3	Diastolic (congestive) heart failure
I50.30	Unspecified diastolic (congestive) heart failure
I50.31	Acute diastolic (congestive) heart failure
I50.32	Chronic diastolic (congestive) heart failure
I50.33	Acute on chronic diastolic (congestive) heart failure
I50.4	Combined systolic (congestive) and diastolic (congestive) heart failure
I50.40	Unspecified combined systolic (congestive) and diastolic (congestive) heart failure
I50.41	Acute combined systolic (congestive) and diastolic (congestive) heart failure
I50.42	Chronic combined systolic (congestive) and diastolic (congestive) heart failure
I50.43	Acute on chronic combined systolic (congestive) and diastolic (congestive) heart failure
I50.9	Heart failure, unspecified
I11.0	Hypertensive heart disease with (congestive) heart failure
I13.0	Hypertensive heart and chronic kidney disease with heart failure and stage 1 through stage 4 chronic kidney disease, or unspecified chronic kidney disease
I13.2	Hypertensive heart and chronic kidney disease with heart failure and with stage 5 chronic kidney disease, or end stage renal disease
I97.13	Postprocedural heart failure
I97.130	Postprocedural heart failure following cardiac surgery
I97.131	Postprocedural heart failure following other surgery

	I09.81	Heart failure rheumatic (chronic) (inactive) (with chorea)

Analytic variables and permanent datasets to be created

Definitions/calculations:

Create an analytic dataset with the following variables:

Variable	Type	Notes
study_id	char(6)	Primary key
height	integer	Height of the patient in cm
member	boolean	Indicator variable for health plan membership
index_date	date	Date of the first time the patient was dispensed either drug
age_at_index	integer	Age as of the index date
drug_a_supply	integer	Sum of days supply where drug A was dispensed
drug_b_supply	integer	Sum of days supply where drug A was dispensed
dx_outcome	boolean	Outcome variable for analysis of the HF outcome. Set to 1 when the patient had an HF dx and 0 when they did not.

Save this dataset to a permanent location to ensure this analysis can be reproduced later if needed.

STATISTICAL ANALYSIS

Logistic regression modeling the risk of a heart failure diagnosis following exposure to either drug A or drug B, adjusted for patient height, age, and health-plan membership.

If time allows, also create the following table summarizing the characteristics of the study cohort, by heart failure outcome, with adjusted P-values for the significance of each explanatory variable in the model:

	Heart Failure Diagnosis			
Membership, n (%)*	Any Heart Failure, N=xx	No Heart Failure, N=xx	Overall, N=xx	P-value
No	xx (xx%)	xx (xx%)	xx (xx%)	0.xxx
Yes	xx (xx%)	xx (xx%)	xx (xx%)	
Height, mean (std)	xx (xx)	xx (xx)	xx (xx)	0.xxx
Age at Index, mean (std)	xx (xx)	xx (xx)	xx (xx)	0.xxx
Days supplied, mean (std)	xx (xx)	xx (xx)	xx (xx)	0.xxx
Drug A, mean (std)	xx (xx)	xx (xx)	xx (xx)	0.xxx
Drug B, mean (std)	xx (xx)	xx (xx)	xx (xx)	0.xxx

*Column percentages