**GWAS Analysis Plan for Hypoglycemia**

**Objective:** To conduct a genome-wide association study (GWAS) of hypoglycemia in those who have been diagnosed with diabetes in the Million Vet Program (MVP) study population. The study will determine the odds of having hypoglycemia given the genotypes at specific loci.

**Data:** In our GWAS, participants are included in the study if they meet **all** these criteria:

* They had one primary care diagnosis of diabetes or 2 or more diagnoses anywhere
* They filled at least one oral or injectable prescription (INS\_SULF\_FLAG)
* They had at least one random plasma glucose (RPG) measurement after diabetes diagnosis
* Their diabetes diagnosis (FIRST\_DIABETES\_ DT) was before enrollment in MVP (ENROLLED\_DATE)
* They have been consistent users of the VA for health care

**Methods:** After the assessment of the inclusion criteria, our study population consists of 114,096 participants. We will conduct the GWAS using a case-control study design. A case is defined as a participant who has hypoglycemia: a participant who received a diagnosis from an Emergency Department (ED) visit (ED\_HYPO\_DX\_FLAG) or an outpatient RPG measure < 70 (GLUCOSE\_LT70\_FLAG) at any point after diabetes diagnosis. A control is defined as a participant that met all the inclusion criteria but did not meet the case definition. Our study will have 38,114 cases and 75,982 controls.

The GWAS study will be performed using a logistic regression model to associate case status (CASE\_CONTROL\_FLAG) with genotype data (what platform was used for genotyping?) for each participant. We will use an additive model of the effect allele (minor allele?): 0, 1, or 2 at each locus as the main predictor in our model. We will adjust the model for age at enrollment (AGE) in MVP, sex (GENDER), and the top 10 principal components of ancestry. The results of the study will be stratified by race. (I wrote this down but not sure it makes sense to me anymore - Do we want to include an interaction term in the model to account for this?)

We will look for significant variants using a Bonferroni-corrected threshold of 5x10-7.

Do we need to calculate the PCs ahead of time or will that be part of the GWAS? Do we need to do any SNP trimming (Hardy-Weinberg, missing SNPs, homozygous SNPs? Outliers of principal components?)

**Secondary Analysis:** As a result of the information gleaned from the GWAS test, we will conduct a matched Case-Control analysis to account for potential confounders that could influence the results. Potential confounders include HbA1c levels, eGFR values, diabetes duration, exposure and duration of prescription type. Ideally, the primary analysis will uncover specific variants that will focus our analysis in the second phase. If specific variants do not arise, the secondary analysis will be conducted again as a GWAS test. An alternative design is to utilize the cases as their own controls by defining a precise before/after definition for the occurrence of the hypoglycemia event.