**Use of Synthea™ to Validate a T2DM Phenotype**

**Abstract**

We generated a synthetic cohort of patients using Synthea™ version 2.4.0. In total, XX patients were generated. Following the implementation of a T2DM phenotype algorithm in Python, we identified XX cases and XX controls.

**Methods**

We downloaded Synthea™ version 2.4.0 from its GitHub repository. Using the default configuration options, we ran it to generate a purely random cohort of individuals. Taking a previously published Type 2 Diabetes Mellitus phenotype definition, we implemented a new version as a Jupyter notebook with the Python 2 kernel. Statistics were calculated using R 3.5.2, and numeric results, tables and figures in this manuscript were generated using StatTag 5.0.0.

**Results**

In total, XX patients were generated by Synthea™. Our implementation of the T2DM phenotype algorithm identified XX cases and XX controls. Additional characteristics for the cases and controls is shown in Table 1, and the counts of encounters for all patients are shown in Figure 1.

**Table 1.** Population characteristics for cases and controls.

**Figure 1.** Counts of encounters for all patients.

**Appendix A – R version**