**Urine Metabolic Signatures Data Guide**

**Overview**

The following R dataframes (.rds) are contained within the directory:

1. ***medn\_cnmc.rds***: contains type 1 diabetes status, patient demographics, and measured metabolite abundances for patients from cohort 1 (CNMC).
2. ***medn\_bdcd.rds***: contains type 1 diabetes status, patient demographics, and measured metabolite abundances for patients from cohort 2 (BDCD).
3. ***medn\_iusom.rds***: contains type 1 diabetes status, patient demographics, and measured metabolite abundances for patients from cohort 3 (IUSOM).

The acronyms CNMC, BDCD, and IUSOM refer to the clinical sites that each cohort were recruited from.

**Dataset details**

All data contain the same set of variables, however the IUSOM data has N/A for all values in the SIBLINGS column given that recruited patients at that site were not sibling case-control pairs. The descriptions of all variable columns are provided below.

1. ***medn\_cnmc.rds***
   1. 64 rows: Each corresponds to a unique sample
   2. 96 columns:
      1. **SUBJECT**: columns containing patient identifiers.
      2. **OUTCOME**: column containing type 1 diabetes status.
      3. **AGE**: column containing patient age in years.
      4. **GENDER**: column containing patient gender.
      5. **SIBLINGS**: column containing unique identifiers for sibling pairs.
      6. **X2.3.dihydroxybutanoic.acid. – xylitol**: columns containing metabolite abundances.
2. ***medn\_bdcd.rds***
   1. 60 rows: Each corresponds to a unique sample
   2. 96 columns:
      1. **SUBJECT**: columns containing patient identifiers.
      2. **OUTCOME**: column containing type 1 diabetes status.
      3. **AGE**: column containing patient age in years.
      4. **GENDER**: column containing patient gender.
      5. **SIBLINGS**: column containing unique identifiers for sibling pairs.
      6. **X2.3.dihydroxybutanoic.acid. – xylitol**: columns containing metabolite abundances.
3. ***medn\_iusom.rds***
   1. 32 rows: Each corresponds to a unique sample
   2. 96 columns:
      1. **SUBJECT**: columns containing patient identifiers.
      2. **OUTCOME**: column containing type 1 diabetes status.
      3. **AGE**: column containing patient age in years.
      4. **GENDER**: column containing patient gender.
      5. **SIBLINGS**: column containing unique identifiers for sibling pairs.
      6. **X2.3.dihydroxybutanoic.acid. – xylitol**: columns containing metabolite abundances.