# Data dictionary for

## Longitudinal analysis reveals high prevalence of Epstein-Barr Virus associated with multiple sclerosis

**This file includes:**

Data dictionaries for datasets

Filename: baseline_chr_all.csv
R program: baseline_characteristics.R
Purpose: Reproduce Fig. 1A

| Variable | Description | Values |
|---|---|---|
| sex | Sex | Categorical: Male; Female. |
| race | Race | Categorical: Black; White; Hispanic; Other. |
| age_baseline | Age at baseline (first sample) | Categorical: <26; 26-30; 31-35; 36-40; >40. |

Filename: baseline_chr_cases.csv
R program: baseline_characteristics.R
Purpose: Reproduce Fig. 1B and 1C

| Variable | Description | Values |
|---|---|---|
| age_baseline | Age at baseline (first sample) | Numeric: Years |
| age_onset | Age at MS onset | Numeric: Years |

Filename: ebv_data.csv
R program: ebv_analyses.R
Purpose: Reproduce Fig. 2, Fig. 3, Fig. S3, Fig. S4, and estimates in the manuscript

| Variable | Description | Values |
|---|---|---|
| id | Random individual ID | Numeric: 1-2387 |
| group | Random ID for matched groups | Numeric: 1-810 |
| serumn | Serum number | 1: First, 2: Second, 3: Third |
| casestat | Case-control status | 1: MS case, 0: Control |
| time_from_bsl | Time since baseline | Numeric: Years |
| time_to_MS | Time to MS onset | Numeric: Years |
| ebv_pos | EBV-status | 1: Positive, 0: Negative |
| cmv_pos | CMV-status | 1: Positive, 0: Negative |
| nfl | Neurofilament light chain | Numeric: pg/ml |
| nfl_log_adj | Age- and sex-adjusted neurofilament light chain | Numeric: log(pg/ml) |

Filename: Virscan_dataset.csv
R program: VirScan_analyses_program.R
Purpose: Reproduce Fig. 4, Table S1, and Table S2.

| Variable | Description | Values |
|---|---|---|
| id1 | Random individual ID | Numeric: 1-60 |
| casestat | Case-control status | 1: MS case, 0: Control |
| timeser | Timing of serum sample | “last”: pre-, “after”: post-onset |
| repnum | Same sample assay replicates | 1: run 1, 2: run 2 |
| pep_id | Viral peptide ID | Numeric |
| zscore | Antibody enrichment Z-score | Numeric |
| Species | Viral species | Character |
| serotype | Viral serotype | Character |
| Organism | Viral strain | Character |
| Protein_name | Viral protein | Character |
| UniProt_acc | UniProt ID | Character |
| pro_len | Viral protein length | Numeric |
| start | Peptide position, start | Numeric |
| end | Peptide position, end | Numeric |