**Transmission of Oral microbiome and Sequencing (TOS) Project**

**Readme - Symptom and Shedding Duration Dataset**

**Datasets**

1. **study\_data.csv**: data on secondary cases of influenza virus infection.

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Description** | **Possible Values** |
| ID | Unique participant ID | P001 to P717 (not inclusive; only a subset is used) |
| SampleID | Unique sample ID | T0001 to T1422 (not inclusive; only a subset is used) |
| comp\* | Bacterial community type | 1 to 5 |
| PCR | Influenza subtype by RT-PCR | 0 = no detection;1 = H1N1;2 = H3N2;3 = B;4 = Co-infection |
| HAI | Influenza subtype by hemagglutination inhibition | 0 = no detection;1 = H1N1;2 = H3N2;3 = B |
| fever | Indicates whether participant had a fever during follow up | 0 = no;1 = yes |
| rhino | Indicates whether participant had rhinorrhea during follow up | 0 = no;1 = yes |
| throat | Indicates whether participant had a sore throat during follow up | 0 = no;1 = yes |
| cough | Indicates whether participant had a cough during follow up | 0 = no;1 = yes |

\*Dirichlet multinomial mixture modeling, an unsupervised clustering method, was used to assign

nose/throat samples to 5 bacterial community types. 114 samples from 124 secondary cases were assigned a community type using a ≥90% posterior probability threshold.

Reference: Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for 529 microbial metagenomics. PloS One. 2012;7: e30126. doi:10.1371/journal.pone.0030126

For a full description of community definitions and methodology: Lee KH, Gordon A, Shedden K, Kuan G, Ng S, Balmaseda A, Foxman B. The respiratory microbiome and susceptibility to influenza virus infection. PloS One. 2019;14:e0207898. doi:10.1371/journal.pone.0207898.

1. **Survival time datasets**: data on duration of symptoms, duration of viral shedding, serial interval, and time to viral shedding (in days). Each dataset includes participants with respective symptom or shedding status.

|  |  |
| --- | --- |
| **Variable Name** | **Description** |
| ID | Unique participant ID |
| comp\* | Bacterial community type |
| time1 | Survival time, lower bound |
| time2 | Survival time, upper bound |

* 1. **fever\_dur.csv:** fever duration
	2. **rhino\_dur.csv:** rhinorrhea duration
	3. **throat\_dur.csv:** sore throat duration
	4. **cough\_dur.csv:** cough duration
	5. **shed\_dur.csv:** viral shedding duration
	6. **sym\_onset.csv:** serial interval
	7. **shed\_onset.csv:** time to viral shedding
1. **oligotype\_count.csv**: oligotype count table. Includes samples from all household members. For a full description of the oligotyping process: Lee KH, Gordon A, Shedden K, Kuan G, Ng S, Balmaseda A, Foxman B. The respiratory microbiome and susceptibility to influenza virus infection. PloS One. 2019;14:e0207898. doi:10.1371/journal.pone.0207898.
2. **tax\_data.csv:** taxonomic classification of oligotypes. Taxonomic classification obtained by searching representative sequences of each oligotype against the Human Oral Microbiome Database (HOMD) v14.51 (http://www.homd.org/) using blastn v2.2.23 (https://blast.ncbi.nlm.nih.gov/Blast.cgi). For a full description of the taxonomy assignment process: Lee KH, Gordon A, Shedden K, Kuan G, Ng S, Balmaseda A, Foxman B. The respiratory microbiome and susceptibility to influenza virus infection. PloS One. 2019;14:e0207898. doi:10.1371/journal.pone.0207898.

**Data citation**: Lee KH, Gordon A, Foxman B. (2019). Transmission of Oral microbiome and Sequencing - Symptom and Shedding Duration [Data set]. University of Michigan Deep Blue Data Repository. <https://doi.org/10.7302/Z2W66J18>

**Citation to related material:**

Lee KH, Gordon A, Shedden K, Kuan G, Ng S, Balmaseda A, Foxman B. The respiratory microbiome and susceptibility to influenza virus infection. PloS One. 2019;14:e0207898. doi:10.1371/journal.pone.0207898.