**Supplemental file 2.** Size of Simplified Lund Covid-19 Corpora. All contain 2419 entities, same as the original Lund Covid-19 Gold Standard Corpus

|  |  |  |  |
| --- | --- | --- | --- |
| **Simplified corpus** | **B-tags** | **I-tags** | **O-tags** |
| Disease | 30 | 33 | 2356 |
| Protein (gene) | 17 | 0 | 2402 |
| Species | 119 | 58 | 2242 |