|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Supplemental Table 1. Counts of features included in the model by category** | | | | | |
| **Feature Category** | | | **UM** | **MGH** |
| **Patient Demographics** | | |  |  |
|  | Age (quintiles) | | 5 | 5 |
|  | Gender | | 1 | 1 |
|  | Race | | 8 | 6 |
|  | Marital status | | 2 | 5 |
|  | County | | 123 | -- |
|  | State | | 19 | -- |
|  | BMI | | 5 | -- |
|  | Insurance | | 6 | -- |
|  | Total demographic features | | **169** | **17** |
| **Admission** | | |  |  |
|  | Admission type | | 4 | 1 |
|  | Patient type | | 4 | -- |
|  | Location unit(/room for UM) | | 333 | 39 |
|  | Admission Service | | -- | 12 |
|  | Admission Source | | -- | 2 |
|  | Total admission features | | **341** | **54** |
| **Vitals** | | | **17** | **--** |
| **Laboratory Tests** | | | **270** | **405** |
| **Procedures** | | | **121** | **122** |
| **Medications** | | |  |  |
|  | Medication | | 866 | 624 |
|  | Ingredient | | 562 | -- |
|  | Class | | 219 | -- |
|  | Total medication features | | **1647** | **624** |
| **Colonization Pressure** | | | **10** | **7** |
| **History** | | |  |  |
|  | History of CDI | | 2 | 2 |
|  | Previous encounters (stats) | | 13 | 9 |
|  | Diagnoses (DRG/ICD9/ICD10) | | 733 | 259 |
|  | Medications | |  |  |
|  |  | Medication | 786 | *338* |
|  |  | Ingredient | 514 | -- |
|  |  | Class | 213 | -- |
|  |  | Total historical medication features | **1513** | **338** |
|  | Total historical features | | **2261** | **608** |
| **Total features (all categories)** | | | **4,836** | **1,837** |

After processing, we had 4,836 features for UM and 1,837 features at MGH. Feature categories are in bold. Subcategories are listed below bolded categories and indented. Some features were not available at both institutions (those marked “--“). In particular, vitals were only available at UM, while MGH had richer laboratory data.

BMI: body mass index; CDI: *Clostridium difficile* infection; DRG: Diagnosis-related group; ICD9: International Classification of Diseases, Ninth Revision; ICD 10: Classification of Diseases, Tenth Revision