

Machine learning and prediction of incident psychiatric hospitalizations using NYC-CDRN data

Project Goal: Prediction of incident psychiatric hospitalization (*bipolar/manic disorders, depression, schizophrenia, psychoses/delusional disorders*), using a 1-year study window

Time period: 01/2013-12/2017

Study Design: Retrospective case-control

Data Source: CDRN [OMOP]

Population:

Potential Cases: 470,242 pts with hx of depression Dx or antidepressant Rx

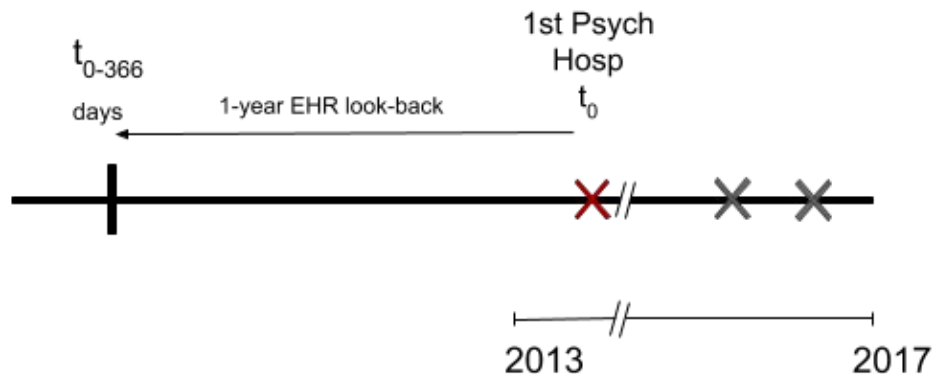
Potential Controls: 1,000,000 pts with no hx of psychiatric Dx or antidepressant Rx

Psychiatric ICD Codes

	ICD-9-CM	ICD-10-CM
All Psychiatric Diagnoses	293.8[1-2], 295, 296, 297, 298, 300.4, 311	F06.[0,2], F20, F21, F22, F23, F24, F25, F28, F29, F30, F31, F32, F33, F34.[1,8], F39
Bipolar / Manic / Other Mood	296.[0,1,4-9]	F30, F31, F34.8, F39
Depression	296.2, 296.3, 300.4, 311	F32, F33, F34.1
Schizophrenia	295	F20, F21, F25
Psychoses (<i>incl.</i> Delusional disorders)	293.8[1-2], 297, 298	F06.[0,2], F22, F23, F24, F28, F29

Case Identification

1. Psych ICD code (**primary, secondary**) + Inpatient Hospitalization, 2013-2017
 - a. Total, $n = 8,896$
 - b. After exclusion, $n = 7,166$
2. Psych ICD code (**primary**) + Inpatient Hospitalization, 2013-2017
 - a. Total, $n = 1,701$
 - b. After exclusion, $n = 1,038$

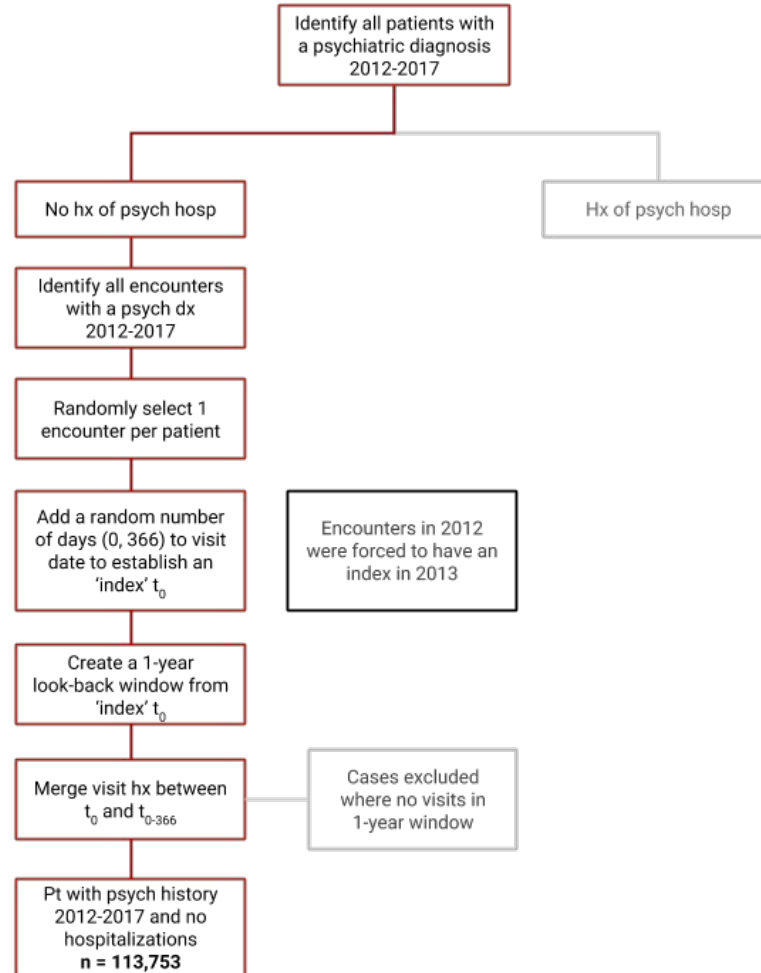


Inclusion criteria:

- Age ≥ 18
- ≥ 2 visits in 1-year window
- ≥ 1 days between first-last visit
- ≥ 1 dx mood | schizophrenia (CCS Categories)

'Control' Identification

- All patients have a psychiatric diagnosis in year leading up to 'index' t_0 .
- Diagnosis encounter and index dates chosen at random



Feature Definitions

CCS Algorithm for Diagnoses (ICD-9)

Appendix A – Clinical Classification Software-DIAGNOSES (January 1980 through September 2015)

Revised 03/24/2016

1	Tuberculosis	01000 01001 01002 01003 01004 01005 01006 01010 01011 01012 01013 01014 01015 01016 01080 01081 01082 01083 01084 01085 01086 01090 01091 01092 01093 01094 01095 01096 01100 01101 01102 01103 01104 01105 01106 01110 01111 01112 01113 01114 01115 01116 01120 01121 01122 01123 01124 01125 01126 01130 01131 01132 01133 01134 01135 01136 01140 01141 01142 01143 01144 01145 01146 01150 01151 01152 01153 01154 01155 01156 01160 01161 01162 01163 01164 01165 01166 01170 01171 01172 01173 01174 01175 01176 01180 01181 01182 01183 01184 01185 01186 01190 01191 01192 01193 01194 01195 01196 01200 01201 01202 01203 01204 01205 01206 01210 01211 01212 01213 01214 01215 01216 01220 01221 01222 01223 01224 01225 01226 01230 01231 01232 01233 01234 01235 01280 01281 01282 01283 01284 01285 01286 01300 01301 01302 01303 01304 01305 01306 01310 01311 01312 01313 01314 01315 01316 01320 01321 01322 01323 01324 01325 01326 01330 01331 01332 01333 01334 01335 01336 01340 01341 01342 01343 01344 01345 01346 01350 01351 01352 01353 01354 01355 01356 01360 01361 01362 01363 01364 01365 01366 01380 01381 01382 01383 01384 01385 01386 01390 01391 01392 01393 01394 01395 01396 01400 01401 01402 01403 01404 01405 01406 01480 01481 01482 01483 01484 01485 01486 01500 01501 01502 01503 01504 01505 01506 01510 01511 01512 01513 01514 01515 01516 01520 01521 01522 01523 01524 01525 01526 01550 01551 01552 01553 01554 01555 01556 01560 01561 01562 01563 01564 01565 01566 01570 01571 01572 01573 01574 01575 01576 01580 01581 01582 01583 01584 01585 01586 01590 01591 01592 01593 01594 01595 01596 01600 01601 01602 01603 01604 01605 01606 01610 01611 01612 01613 01614 01615 01616 01620 01621 01622 01623 01624 01625 01626 01630 01631 01632 01633 01634 01635 01636 01640 01641 01642 01643 01644 01645 01646 01650 01651 01652 01653 01654 01655 01656 01660 01661 01662 01663 01664 01665 01666 01670 01671 01672 01673 01674 01675 01676 01690 01691 01692 01693 01694 01695 01696 01700 01701 01702 01703 01704 01705 01706 01710 01711 01712 01713 01714 01715 01716 01720 01721 01722 01723 01724 01725 01726 01730 01731 01732 01733 01734 01735 01736 01740 01741 01742 01743 01744 01745 01746 01750 01751 01752 01753 01754 01755 01756 01760 01761 01762 01763 01764 01765 01766 01770 01771 01772 01773 01774 01775 01776 01780 01781 01782 01783 01784 01785 01786 01790 01791 01792 01793 01794 01795 01796 01800 01801 01802 01803 01804 01805 01806 01880 01881 01882 01883 01884 01885 01886 01890 01891 01892 01893 01894 01895 01896 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2818 2819 2820 2821 2822 2823 2824 2825 2826 2827 2828 2829 2830 2831 2832 2833 2834 2835 2836 2837 2838 2839 2840 2841 2842 2843 2844 2845 2846 2847 2848 2849 2850 2851 2852 2853 2854 2855 2856 2857 2858 2859 2860 2861 2862 2863 2864 2865 2866 2867 2868 2869 2870 2871 2872 2873 2874 2875 2876 2877 2878 2879 2880 2881 2882 2883 2884 2885 2886 2887 2888 2889 2890 2891 2892 2893 2894 2895 2896 2897 2898 2899 2900 2901 2902 2903 2904 2905 2906 2907 2908 2909 2910 2911 2912 2913 2914 2915 2916 2917 2918 2919 2920 2921 2922 2923 2924 2925 2926 2927 2928 2929 2930 2931 2932 2933 2934 2935 2936 2937 2938 2939 2940 2941 2942 2943 2944 2945 2946 2947 2948 2949 2950 2951 2952 2953 2954 2955 2956 2957 2958 2959 2960 2961 2962 2963 2964 2965 2966 2967 2968 2969 2970 2971 2972 2973 2974 2975 2976 2977 2978 2979 2980 2981 2982 2983 2984 2985 2986 2987 2988 2989 2990 2991 2992 2993 2994 2995 2996 2997 2998 2999 3000 3001 3002 3003 3004 3005 3006 3007 3008 3009 3010 3011 3012 3013 3014 3015 3016 3017 3018 3019 3020 3021 3022 3023 3024 3025 3026 3027 3028 3029 3030 3031 3032 3033 3034 3035 3036 3037 3038 3039 3040 3041 3042 3043 3044 3045 3046 3047 3048 3049 3050 3051 3052 3053 3054 3055 3056 3057 3058 3059 3060 3061 3062 3063 3064 3065 3066 3067 3068 3069 3070 3071 3072 3073 3074 3075 3076 3077 3078 3079 3080 3081 3082 3083 3084 3085 3086 3087 3088 3089 3090 3091 3092 3093 3094 3095 3096 3097 3098 3099 3100 3101 3102 3103 3104 3105 3106 3107 3108 3109 3110 3111 3112 3113 3114 3115 3116 3117 3118 3119 3120 3121 3122 3123 3124 3125 3126 3127 3128 3129 3130 3131 3132 3133 3134 3135 3136 3137 3138 3139 3140 3141 3142 3143 3144 3145 3146 3147 3148 3149 3150 3151 3152 3153 3154 3155 3156 3157 3158 3159 3160 3161 3162 3163 3164 3165 3166 3167 3168 3169 3170 3171 3172 3173 3174 3175 3176 3177 3178 3179 3180 3181 3182 3183 3184 3185 3186 3187 3188 3189 3190 3191 3192 3193 3194 3195 3196 3197 3198 3199 3200 3201 3202 3203 3204 3205 3206 3207 3208 3209 3210 3211 3212 3213 3214 3215 3216 3217 3218 3219 3220 3221 3222 3223 3224 3225 3226 3227 3228 3229 3230 3231 3232 3233 3234 3235 3236 3237 3238 3239 3240 3241 3242 3243 3244 3245 3246 3247 3248 3249 3250 3251 3252 3253 3254 3255 3256 3257 3258 3259 3260 3261 3262 3263 3264 3265 3266 3267 3268 3269 3270 3271 3272 3273 3274 3275 3276 3277 3278 3279 3280 3281 3282 3283 3284 3285 3286 3287 3288 3289 3290 3291 3292 3293 3294 3295 3296 3
---	--------------	--

Analysis Plan

Construct several predictive models using the below statistical learning methods.

Compare confusion matrix, classification report, AUC score on:

1. Logistic Regression (LR)
 - a. L1 regularization
 - b. L2 regularization
2. Random Forest (RF)
3. Extreme Gradient Boosting (XGB)
4. Multi-layer Perceptron (MLP)

Model 1:

- **Primary, Secondary** Dx Hospitalization (outcome)
- All samples (1:X, case:control)

Model 2:

- **Primary, Secondary** Dx Hospitalization (outcome)
- Random subsample controls (1:1, case:control)

Model 3:

- **Primary, Secondary** Dx Hospitalization (outcome)
- SMOTE (1:1 ratio, case:control)

Analysis Plan

Construct several predictive models using the below statistical learning methods.

Compare confusion matrix, classification report, AUC score on:

1. Logistic Regression (LR)
 - a. L1 regularization
 - b. L2 regularization
2. Random Forest (RF)
3. Extreme Gradient Boosting (XGB)
4. Multi-layer Perceptron (MLP)

Model 4:

- **Primary** Dx Hospitalization (outcome)
- All samples (1:X, case:control)

Model 5:

- **Primary** Dx Hospitalization (outcome)
- Random subsample controls (1:1, case:control)

Model 6:

- **Primary** Dx Hospitalization (outcome)
- SMOTE (1:1 ratio, case:control)

Model 1: Results

- Primary, secondary Dx Hospitalization
- All samples (1:X, case:control)
 - $n_{\text{case}} = 7,166$; $n_{\text{control}} = 89,716$

Logistic Regression

Confusion Matrix – Test			
TN	22452	4418	FP
FN	720	1475	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.97	0.84	0.90
1	0.25	0.67	0.36

AUC Score
0.754

Random Forest

Confusion Matrix – Test			
TN	25555	1315	FP
FN	1299	896	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.95	0.95	0.95
1	0.41	0.41	0.41

AUC Score
0.680

XGBoost

Confusion Matrix – Test			
TN	22422	4448	FP
FN	600	1595	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.97	0.83	0.90
1	0.26	0.73	0.39

AUC Score
0.781

Model 2: Results

- Primary, secondary Dx Hospitalization
- Random subsample controls (1:1 ratio, case:control)
 - $n_{total} = 14,332$

Logistic Regression

Confusion Matrix – Test			
TN	22002	4868	FP
FN	662	1533	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.97	0.82	0.89
1	0.24	0.70	0.36

AUC Score
0.757

Random Forest

Confusion Matrix – Test			
TN	20972	5898	FP
FN	554	1641	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.97	0.78	0.87
1	0.22	0.75	0.34

AUC Score
0.764

XGBoost

Confusion Matrix – Test			
TN	21603	5267	FP
FN	554	1641	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.97	0.80	0.88
1	0.24	0.75	0.36

AUC Score
0.776

Model 3: Results

- Primary, secondary Dx Hospitalization
- SMOTE (1:1 ratio, case:control)
 - $n_{total} = 125,692$

Logistic Regression

Confusion Matrix – Test			
TN	21979	4891	FP
FN	732	1463	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.97	0.82	0.89
1	0.23	0.67	0.34

AUC Score
0.742

Random Forest

Confusion Matrix – Test			
TN	26563	307	FP
FN	1681	514	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.94	0.99	0.96
1	0.63	0.23	0.34

AUC Score
0.611

XGBoost

Confusion Matrix – Test			
TN	26629	241	FP
FN	1609	586	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.94	0.99	0.97
1	0.71	0.27	0.39

AUC Score
0.629

Model 1,2,3: Multilayer Perceptron Results

- Primary, secondary Dx Hospitalization

MLP All Samples

Confusion Matrix – Test			
TN	25830	1040	FP
FN	1487	708	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.95	0.96	0.95
1	0.41	0.32	0.36

AUC Score
0.641

MLP Subsample

Confusion Matrix – Test			
TN	19909	6961	FP
FN	562	1603	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.97	0.74	0.84
1	0.19	0.73	0.30

AUC Score
0.736

MLP SMOTE

Confusion Matrix – Test			
TN	24377	2493	FP
FN	1247	948	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.95	0.91	0.93
1	0.28	0.43	0.34

AUC Score
0.670

Model 4: Results

- Primary Dx Hospitalization
- All samples (1:X, case:control)
 - $n_{\text{case}} = 1,038$; $n_{\text{control}} = 89,716$

Logistic Regression

Confusion Matrix – Test			
TN	23552	3376	FP
FN	84	215	TP

Classification Report – Test			
	Precision	Recall	F1
0	1.00	0.87	0.93
1	0.06	0.72	0.11

AUC Score
0.797

Random Forest

Confusion Matrix – Test			
TN	26915	13	FP
FN	261	38	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.99	1.00	0.99
1	0.75	0.13	0.22

AUC Score
0.563

XGBoost

Confusion Matrix – Test			
TN	23088	3840	FP
FN	63	236	TP

Classification Report – Test			
	Precision	Recall	F1
0	1.00	0.86	0.92
1	0.06	0.79	0.11

AUC Score
0.823

Model 5: Results

- Primary Dx Hospitalization
- Random subsample controls (1:1 ratio, case:control)
 - $n_{total} = 2,076$

Logistic Regression

Confusion Matrix – Test			
TN	21750	5178	FP
FN	73	226	TP

Classification Report – Test			
	Precision	Recall	F1
0	1.00	0.81	0.89
1	0.04	0.76	0.08

AUC Score
0.782

Random Forest

Confusion Matrix – Test			
TN	22910	4018	FP
FN	64	235	TP

Classification Report – Test			
	Precision	Recall	F1
0	1.00	0.85	0.92
1	0.06	0.79	0.10

AUC Score
0.818

XGBoost

Confusion Matrix – Test			
TN	22606	4322	FP
FN	49	250	TP

Classification Report – Test			
	Precision	Recall	F1
0	1.00	0.84	0.91
1	0.05	0.84	0.10

AUC Score
0.838

Model 6: Results

- Primary Dx Hospitalization
- SMOTE (1:1 ratio, case:control)
 - $n_{total} = 125,576$

Logistic Regression

Confusion Matrix – Test			
TN	23548	3380	FP
FN	100	199	TP

Classification Report – Test			
	Precision	Recall	F1
0	1.00	0.87	0.93
1	0.06	0.67	0.10

AUC Score
0.770

Random Forest

Confusion Matrix – Test			
TN	26903	25	FP
FN	273	26	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.99	1.00	0.99
1	0.51	0.09	0.15

AUC Score
0.543

XGBoost

Confusion Matrix – Test			
TN	26899	29	FP
FN	244	55	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.99	1.00	0.99
1	0.65	0.18	0.29

AUC Score
0.591

Analysis Plan

Construct several predictive models using the below statistical learning methods.

Compare confusion matrix, classification report, AUC score on:

1. Logistic Regression (LR)
 - a. L1 regularization
 - b. L2 regularization
2. Random Forest (RF)
3. Extreme Gradient Boosting (XGB)
4. Multi-layer Perceptron (MLP)

Model 1:

- **Primary, secondary** Dx Hospitalization (outcome)
- Full data set (1:X, case:control)

Model 2:

- **Primary** Dx Hospitalization (outcome)
- Full data set (1:X, case:control)

Model 3:

- **Primary, secondary** Dx Hospitalization (outcome)
- Full Subsample (1:1 ratio, case:control)

Model 4:

- **Primary** Dx Hospitalization (outcome)
- Full Subsample (1:1 ratio, case:control)

Model 1: Results

- Primary, secondary Dx Hospitalization
- Full data set (1:X, case:control)
 - $n_{\text{case}} = 8,896$; $n_{\text{control}} = 113,753$

Logistic Regression

Confusion Matrix – Test			
TN	28638	5531	FP
FN	795	1831	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.97	0.84	0.90
1	0.25	0.70	0.37

AUC Score
0.768

Random Forest

Confusion Matrix – Test			
TN	33946	223	FP
FN	1879	747	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.95	0.99	0.97
1	0.77	0.28	0.42

AUC Score
0.639

XGBoost

Confusion Matrix – Test			
TN	28397	5772	FP
FN	627	1999	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.98	0.83	0.90
1	0.26	0.76	0.38

AUC Score
0.796

Model 2 Results

- Primary Dx Hospitalization
- Full data set (1:X, case:control)
 - $n_{\text{case}} = 1,701$; $n_{\text{control}} = 113,753$

Logistic Regression

Confusion Matrix – Test			
TN	26656	5187	FP
FN	810	1696	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.97	0.84	0.90
1	0.25	0.68	0.36

AUC Score
0.757

Random Forest

Confusion Matrix – Test			
TN	34121	29	FP
FN	351	136	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.99	1.00	0.99
1	0.82	0.28	0.42

AUC Score
0.639

XGBoost

Confusion Matrix – Test			
TN	29580	4570	FP
FN	86	401	TP

Classification Report – Test			
	Precision	Recall	F1
0	1.00	0.87	0.93
1	0.08	0.82	0.15

AUC Score
0.845

Model 3 Results

- Primary, Secondary Dx Hospitalization
- Full Subsample (1:1, case:control)
 - $n_{total} = 17,792$

Logistic Regression

Confusion Matrix – Test			
TN	2197	482	FP
FN	786	1873	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.74	0.82	0.78
1	0.80	0.70	0.75

AUC Score
0.762

Random Forest

Confusion Matrix – Test			
TN	2107	572	FP
FN	496	2163	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.81	0.79	0.80
1	0.79	0.81	0.80

AUC Score
0.800

XGBoost

Confusion Matrix – Test			
TN	2138	541	FP
FN	544	2115	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.80	0.80	0.80
1	0.80	0.80	0.80

AUC Score
0.797

Model 4 Results

- Primary Dx Hospitalization
- Full Subsample (1:1, case:control)
 - $n_{total} = 3,402$

Logistic Regression

Confusion Matrix – Test			
TN	410	86	FP
FN	143	382	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.74	0.83	0.78
1	0.82	0.73	0.77

AUC Score
0.777

Random Forest

Confusion Matrix – Test			
TN	427	69	FP
FN	100	425	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.81	0.86	0.83
1	0.86	0.81	0.83

AUC Score
0.835

XGBoost

Confusion Matrix – Test			
TN	420	76	FP
FN	99	426	TP

Classification Report – Test			
	Precision	Recall	F1
0	0.81	0.85	0.83
1	0.85	0.81	0.83

AUC Score
0.829