

Predicting antibiotic usage for improved healthcare-associated infection risk assessment

Laila Butler-Mills and Jeffrey Wang

Advisor: Daniel K. Sewell

July 18, 2024

Healthcare-Associated Infections (HAIs)

- Infections which develop while a patient is receiving care for a separate condition
 - Can be caused by pathogens resistant to antibiotic treatments
 - Can also be caused by antibiotics suppressing probiotics
- Estimated to affect 1 in 31 hospital patients and 1 in 43 nursing home residents
- ~687,000 HAI cases and ~72,000 deaths in 2015 in the US
- Direct medical costs of at least \$28.4 billion each year in the US

Conclusions

Clostridioides difficile (C. diff)

- C. diff lives in healthy people's intestines, but is typically held in check by "good" bacteria there
- Antibiotics taken for another infection can kill good bacteria and allow C. diff to spread unchecked
- C. diff can release toxins in the colon that damage tissues









Antibiotics

- Previous research identified 9 antibiotics that are linked to a greater risk of C. diff infection
 - Clindamycin
 - o Cefixime
 - Cefdinir
 - Cefuroxime
 - $_{\odot}$ Cefpodoxime
 - Amoxicillin/Clavulanate
 - Moxifloxacin
 - Cefaclor
 - Linezolid



Research Objective

Since antibiotics are a major cause of C. diff infections, it is essential to understand how healthcare facilities prescribe antibiotics in order to effectively model C. diff incidence.

Our research aims to use machine learning techniques to build accurate models for predicting antibiotic usage based on patient data such as diagnoses, age, and other variables.



Conclusions

Premier Health Data

- Collected from U.S. hospital discharge records and billing data
- Models use:
 - Demographics (sex, race, age, etc.)
 - Disease statuses
 - Length of stay
 - Medications
 - Procedures

We are only working with a subset of the Premier Health Data:

- Patients who had at least one inpatient visit related to
 - Urinary tract infection (UTI)
 - \circ Pneumonia
- Collect all other visits from these patients

- Ranges from 2001 to 2021
- ~45 million hospitalization events in sample



Conclusions

Premier Health Data Relations

Tables highlighted in red are used in our analysis.

PATDEMO: patient demographics and general hospital characteristics

PATICD_PROC: ICD procedure codes

lu_icdcode: lookup table with procedural ICD codes

lu_std_payor: lookup for payer types

Diagnosis Imbalance Problem

- Diagnostic related groups (DRGs) are numeric codes classifying patient diagnoses and hospitalization costs
 - $_{\odot}$ Ex: DRG 689 "kidney and urinary tract infections with major complication or comorbidity"
- DRGs in sample are heavily related to UTIs and pneumonia
 - General model may be biased towards these conditions and their associated antibiotics
- Solution is to group data based on each individual DRG
 Create a separate model for each DRG based on all other predictor variables

(Fetter, 1991; Centers for Medicare & Medicaid Services, 2020)

Class Rarity Problem

 Some DRGs have very few patients taking a certain antibiotic within them (i.e. <1% of patients)

 $_{\odot}$ Should we still fit an entire model on them?

 Decide to treat such cases as if we fitted a model that always predicted "no antibiotic" for every input
 Trivially has an accuracy of ≥99%





Conclusions

Decision Trees

- Machine learning technique that generates sequential decision rules which assign data into classes
- Rules are constructed by determining which splits result in the largest gain of some information metric (e.g. Gini impurity)
- Can suffer from overfitting



Conclusions

Random Forests

- Generate unique decision trees and aggregate across each output to determine overall result
 - Generated on different subsets of data and features
- Aggregation helps to balance out any individual tree's overfitting and generates more accurate results





Background Data Method

Results

Conclusions

Area Under Curve (AUC)

- ROC curve measures the true vs. false positive rate
- Models with an ROC curve close to the upper left corner (area under curve is close to 1) are favored
- >0.7 AUC generally regarded as indicating a good model





Predicting ABX Usage for Improved HAI Risk Assessment

Area Under the ROC Curve Statistics

Grouped by Antibiotic

Antibiotic	Mean	Standard Deviation	Minimum	Maximum	Percent of Models With AUC > 0.70
Amoxicillin Clavulanate	0.610	0.045	0.460	0.780	3.155
Cefaclor	NA	NA	NA	NA	NA
Cefdinir	0.708	0.057	0.519	0.818	56.757
Cefixime	NA	NA	NA	NA	NA
Cefpodoxime	0.731	0.102	0.615	0.838	50.000
Cefuroxime	0.627	0.057	0.493	0.827	9.040
Clindamycin	0.641	0.056	0.509	0.879	12.317
Linezolid	0.717	0.054	0.589	0.848	64.615
Moxifloxacin	0.768	0.041	0.637	0.866	93.035



Conclusions

Interactive Plot







Performance

- ~68% of DRG-antibiotic combinations had less than 1% of patients taking the antibiotic
- Around 32% of the remaining constructed models exceeded the 0.7 threshold for ROC AUC
- Overall, this means that ~78% of the DRG-antibiotic models performed well
- Models predicted well for certain antibiotics, such as moxifloxacin and linezolid
- Models were worse at predicting other antibiotics, such as amoxicillin/clavulanate

Performance (cont.)



10 Most Important Predictor Variables for Moxifloxacin





Conclusions

• Model results are inconsistent across antibiotic classes

 Can predict prescription of some antibiotics well, which can help us better understand C. diff infection patterns for some antibiotics

• Potential uses:

- Improve models of regional C. diff spread
- $_{\odot}$ Better assess healthcare facilities' risk of getting C. diff

• Further research:

Improve the models through methods like SMOTE resampling
 Include more predictor variables



Predicting ABX Usage for Improved HAI Risk Assessment

Questions?

References

CDC. (2021, June 21). Health Topics - Healthcare Associated Infections. CDC. Retrieved July 16, 2024,

from https://www.cdc.gov/policy/polaris/healthtopics/hai/index.html

Centers for Medicare & Medicaid Services. (2020). ICD-10-CM/PCS MS-DRG v37.0 Definitions Manual. ICD-10-CM/PCS MS-DRG v37.0 Definitions

Manual. Retrieved July 17, 2024, from https://www.cms.gov/icd10m/version372-fullcode-cms/fullcode_cms/P0371.html

Fetter, R. (1991). Diagnosis Related Groups: Understanding Hospital Performance. Informs Journal on Applied Analytics, 21(1), 6-26.

https://doi.org/10.1287/inte.21.1.6

Gunay, D. (2023). [Random Forest Graphic]. Medium. https://medium.com/@denizgunay/random-forest-af5bde5d7e1e

- Kanade, V. (2022). [Example Decision Tree on Buying a Car]. Spiceworks. <u>https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-decision-tree/</u>
- Kılıç, İ. (2023). [ROC Curve and AUC Graphic]. Medium. https://medium.com/@ilyurek/roc-curve-and-auc-evaluating-model-performance-c2178008b02
- Kon, K. (2019). Clostridium difficile bacteria isolated on black background, 3D illustration [Shutterstock]. News Medical. <u>https://www.news-medical.net/health/What-is-Clostridium-difficile.aspx</u>
- Mayo Clinic Staff. (2023, September 1). *C. difficile infection Symptoms and Causes*. Mayo Clinic. Retrieved July 16, 2024, from https://www.mayoclinic.org/diseases-conditions/c-difficile/symptoms-causes/syc-20351691
- Miller, A., Arakkal, A., & Sewell, D. (2023). Comparison of Different Antibiotics and the Risk for Community-Associated Clostridioides difficile Infection: A Case-Control Study. *Open Forum Infectious Diseases*. <u>https://doi.org/10.1093/ofid/ofad413</u>



Predicting ABX Usage for Improved HAI Risk Assessment



Predicting ABX Usage for Improved HAI Risk Assessment

Thank you!

ISIB Program sponsored by the National Heart Lung and Blood Institute (NHLBI), grant # **HL161716-01**.

Laila Butler-Mills laila-butler-mills@uiowa.edu

Jeffrey Wang jeffrey-wang-1@uiowa.edu