Windows into AMR innovation: Visualizing Innovation in Antibiotics

Presenter:
Venkatesh (Venki) Subramanian[1]
Final Project for Data Visualization (EPPS 6356)

[1]Public Policy and Political Economy,
The University of Texas at Dallas
venkatesh.subramanian@utdallas.edu
https://venkiverse.com

December 3, 2025

See https://amrwindows.venkiverse.com

A reminder on my proposal

Topic

Visualize **Innovation in Antibiotics** at a molecular-class level

Research statement

How much worser are the product market outputs of antibiotic patents than comparable patents with similar inputs?

Method

- Mirror the grammar of Bacterial Antimicrobial Resistance (AMR) burden visualization to visualize antibiotic innovation
- Use granular data at an individual patent molecule level
- Prioritize reproducibility by using publicly available datasets, and open-source tools

The policy problem of Antibiotic Innovation

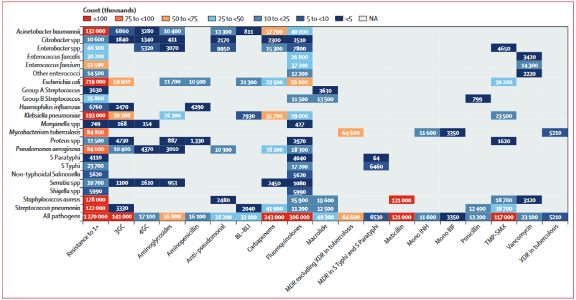


Figure 6: **Global deaths (counts) attributable to bacterial antimicrobial resistance by pathogen-drug combination, 2019**
For this figure, only deaths attributable to resistance, not deaths associated with resistance, are shown due to the very high levels of correlation for resistance patterns between some drugs. 3GC=third-generation cephalosporins. 4GC=fourth-generation cephalosporins. Anti-pseudomonal=anti-pseudomonal penicillin or beta-lactamase inhibitors. BL-BLI=β-lactam or β-lactamase inhibitors. MDR=multidrug resistance. Mono INH=isoniazid mono-resistance. Mono RIF=rifampicin mono-resistance. NA=not applicable. Resistance to 1+=resistance to one or more drug. S Paratyphi=*Salmonella enterica* serotype Paratyphi. S Typhi=*S enterica* serotype Typhi. TMP-SMX=trimethoprim-sulfamethoxazole. XDR=extensive drug resistance.

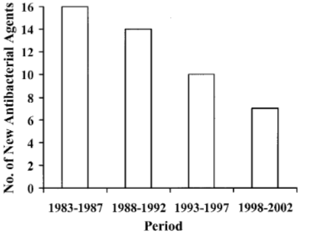


Figure 1. New antibacterial agents approved in the United States, 1983–2002, per 5-year period.

Bacteria are evolving... (Murray et al. 2022)

... but antibiotics are not! (Spellberg et al. 2004)

Innovation in antibiotics lags behind the burden.

Policy solutions are needed to incentivize this innovation.

The visualization problem of Antibiotic Innovation

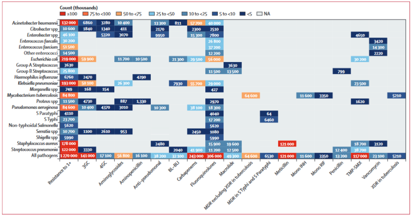


***Figure 6:* Global deaths (counts) attributable to bacterial antimicrobial resistance by pathogen–drug combination, 2019**
For this figure, only deaths attributable to resistance, not deaths associated with resistance, are shown due to the very high levels of correlation for resistance patterns between some drugs. 3GC=third-generation cephalosporins. 4GC=fourth-generation cephalosporins. Anti-pseudomonal=anti-pseudomonal penicillin or beta-lactamase inhibitors. BL-BLI=β-lactam or β-lactamase inhibitors. MDR=multidrug resistance. Mono INH=isoniazid mono-resistance. Mono RIF=rifampicin mono-resistance. NA=not applicable. Resistance to 1+=resistance to one or more drug. S Paratyphi=*Salmonella enterica* serotype Paratyphi. S Typhi=S *enterica* serotype Typhi. TMP-SMX=trimethoprim-sulfamethoxazole. XDR=extensive drug resistance.

(Murray et al. 2022)

DATA: drug = *cat*(`Drugs`)
DATA: patho = *cat*(`Pathogens`)
TRANS: mort = *summary.count*(`2019 AMR Deaths`)
TRANS: mortcol = *cat*(`mort`, *values*(*“>=100”, “75 to <100”, ...*))
ELEMENT: *polygon*(*position*(*bin.rect*(drug*patho)), *color.hue*(mortcol), *label*(mort))

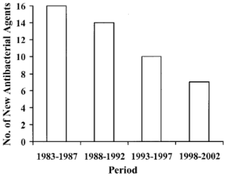


Figure 1. New antibacterial agents approved in the United States, 1983–2002, per 5-year period.

(Spellberg et al. 2004)

DATA: hdec = *cat*(`Half decade`, *values*(*“1983-1987”, “1988-1992”, “1993-1997”, “1998-2002”*))
DATA: inno = `Approved Antibiotics`
ELEMENT: *interval*(*position*(*summary.count*(*bin.rect*(`hdec*inno`))))

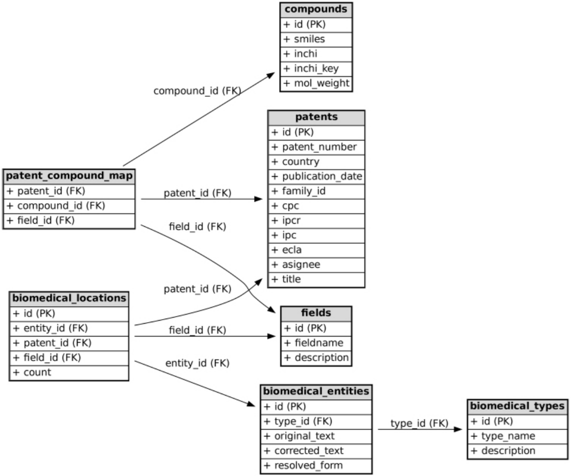
These two graphs must have the same grammar!
(i.e) drug-pathogen level innovation measurement is required.

Datasets

Two publicly available datasets:

- SureChEMBL (Papadatos et al. 2016)
  - Uses ML to annotate the molecules mentioned in worldwide patents
- AntibioticDB (Farrell et al. 2018)
  - A curated list of drug molecules with (potential) antibiotic action

SureChEMBL (Papadatos et al. 2016) I



SureChEMBL (Papadatos et al. 2016) II

Table: Counts of entities in the SureChEMBL database

Patents count	Compounds count	Fields count
43,408,329	30,801,200	6

Note: Database was downloaded on Sep 25, 2025. Patents are counted in each country in which it was granted.

SureChEMBL (Papadatos et al. 2016) III

Table: Counts of compounds by patent fields in the SureChEMBL database

Field description	**Compounds count**
Description	19,117,419
Image (for patents after 2007)	12,825,050
MOL attachments (for patents after 2007)	6,184,796
Claims	5,437,454
Abstract	501,861
Title	162,433

Note: Database was downloaded on Sep 25, 2025.

SureChEMBL (Papadatos et al. 2016) IV

Table: Country-wise patent counts in SureChEMBL

Country	**Patent counts**
CN	22,997,272
US	9,333,627
EP	5,082,085
JP	3,059,911
WO	2,935,432
GB	2
Total	43,408,329

Note: Database was downloaded on Sep 25, 2025.

AntibioticDB (Farrell et al. 2018) I

- It is necessary for the drug names in AntibioticDB and the drug names in Murray et al. (2022) to match. To achieve this I used ChatGPT.
- "You act as a multi-class classifier of antibiotic drug classes. As inputs, you take the drug name, main source describing it, and an untidy drug class name from AntibioticDB in pipe-delimited form. As output, you map the input to one or more or none of the following tidy antibiotic drug class names and provide a brief and concise justification of the classification:
  Third generation cephalosporins, Fourth generation cephalosporins, Aminoglycosides, Aminopenicillin, Anti-pseudomonal penicillin or beta-lactamase inhibitors, Beta-lactam or beta-lactamase inhibitors, Carbapenems, Fluoroquinolones, Macrolide, Multidrug resistance excluding extensive drug resistance in tuberculosis, Multidrug resistance in S Typhi and S Paratyphi, Meticillin, Isoniazid mono-resistance, Rifampicin mono-resistance, Penicillin, Trimethoprim-sulfamethoxazole, Vancomycin, Extensive drug resistance in tuberculosis
  Ensure your output is pipe delimited.
  Examples:
  Input: Amoxicillin — https://www.ncbi.nlm.nih.gov/books/NBK482250/ — Beta-lactam (aminopenicillin)
  Output: Aminopenicillin — Beta lactum is too broad. Aminopenicillin is the most granular category. Amoxicillin is also known as aminopenicillin.
  Input: Cefepime + enmetazobactam (Exblifep) — http://allecra.com/pipeline-2/ — Beta-lactam (cephalosporin, fourth generation) + beta-lactamase inhibitor (penicillanic acid sulfone)
  Output: Beta-lactam or beta-lactamase inhibitors — Beta lactum is too broad. This is more specifically a beta-lactamase inhibitor."

Tools

Reproducibility is the main criterion for tool choice here.

- ggplot2
- Shiny
- Git & GitHub

Conclusion

Topic

Visualize **Innovation in Antibiotics** at a molecular-class level

Research statement

How much worser are the product market outputs of antibiotic patents than comparable patents with similar inputs?

Method

- Mirror the grammar of Bacterial Antimicrobial Resistance (AMR) burden visualization to visualize antibiotic innovation
- Use granular data at an individual patent molecule level
- Prioritize reproducibility by using publicly available datasets, and open-source tools

References I

Farrell, L J, R Lo, J J Wanford, A Jenkins, A Maxwell, and L J V Piddock. 2018. Revitalizing the drug pipeline: AntibioticDB, an open access database to aid antibacterial research and development. *Journal of Antimicrobial Chemotherapy* 73, no. 9 (September): 2284–2297. ISSN: 0305-7453, accessed October 14, 2025. https://doi.org/10.1093/jac/dky208. https://doi.org/10.1093/jac/dky208.

References II

Murray, Christopher J. L., Kevin Shunji Ikuta, Fablina Sharara, Lucien Swetschinski, Gisela Robles Aguilar, Authia Gray, Chieh Han, et al. 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis [in English]. Publisher: Elsevier, *The Lancet* 399, no. 10325 (February): 629–655. ISSN: 0140-6736, 1474-547X, accessed November 25, 2023. https://doi.org/10.1016/S0140-6736(21)02724-0. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)02724-0/fulltext.

References III

- Papadatos, George, Mark Davies, Nathan Dedman, Jon Chambers, Anna Gaulton, James Siddle, Richard Koks, et al. 2016. SureChEMBL: a large-scale, chemically annotated patent document database. *Nucleic Acids Research* 44, no. D1 (January): D1220–D1228. ISSN: 0305-1048, accessed October 8, 2025. https://doi.org/10.1093/nar/gkv1253. https://doi.org/10.1093/nar/gkv1253.
- Spellberg, Brad, John H. Powers, Eric P. Brass, Loren G. Miller, and John E. Edwards Jr. 2004. Trends in Antimicrobial Drug Development: Implications for the Future. *Clinical Infectious Diseases* 38, no. 9 (May): 1279–1286. ISSN: 1058-4838, accessed July 16, 2024. https://doi.org/10.1086/420937. https://doi.org/10.1086/420937.