Visualizing Innovation in Antibiotics

Presenter:
Venkatesh (Venki) Subramanian[1]

Project Proposal for Data Visualization (EPPS 6356)

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

October 15, 2025

Project 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

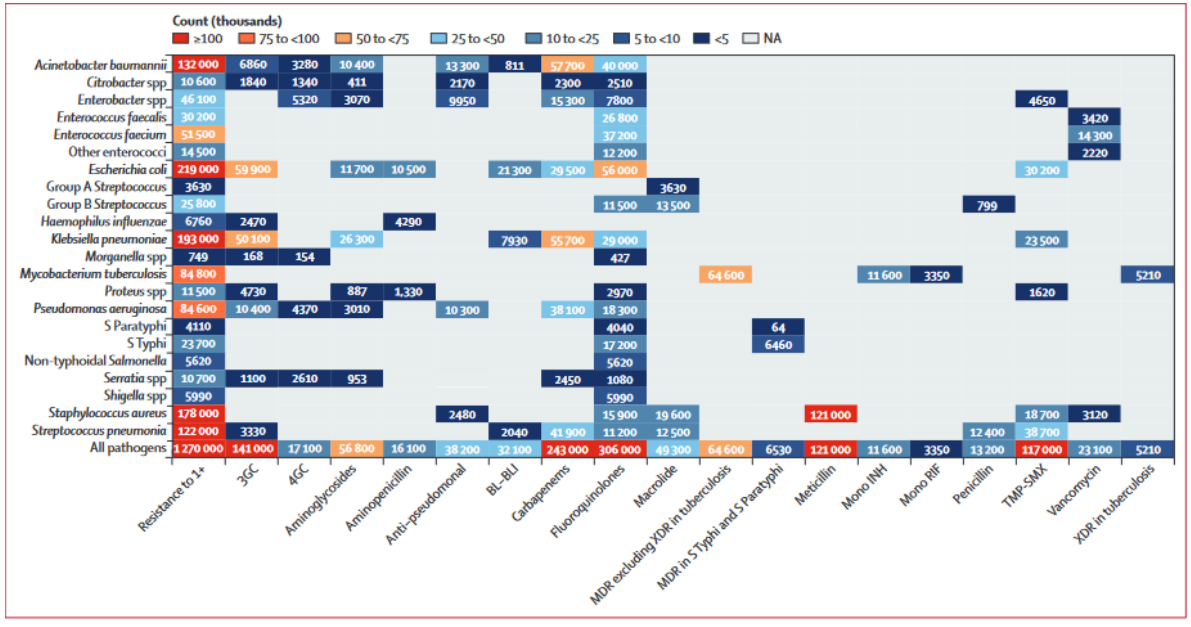

Figure 6: **Global deaths (counts) attributable to bacterial antimicrobial resistance by pathogen-drug combination, 2019**
For this figure, only deaths attributable to resistance, not deaths associated with resistance, are shown due to the very high levels of correlation for resistance patterns between 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.

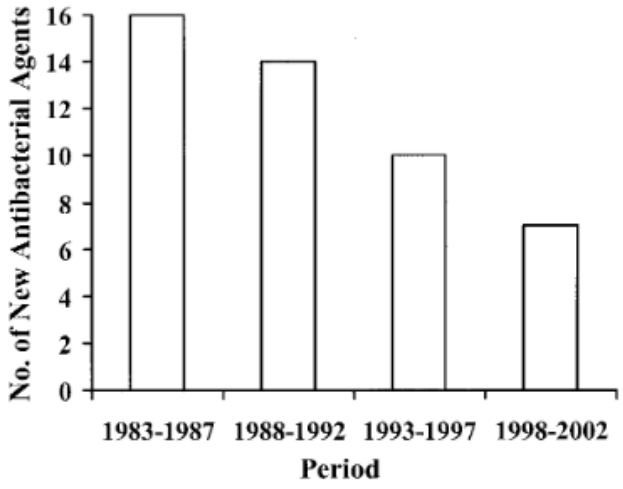

Figure 1. New antibacterial agents approved in the United States, 1983–2002, per 5-year period.

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

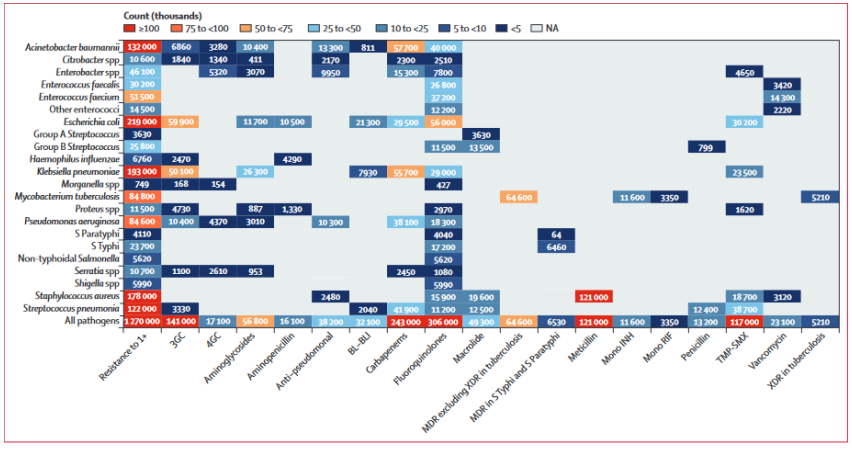

Figure 6: Global deaths (counts) attributable to bacterial antimicrobial resistance by pathogen-drug combination, 2019
For this figure, only deaths attributable to resistance, not deaths associated with resistance, are shown due to the very high levels of correlation for resistance patterns between some 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))

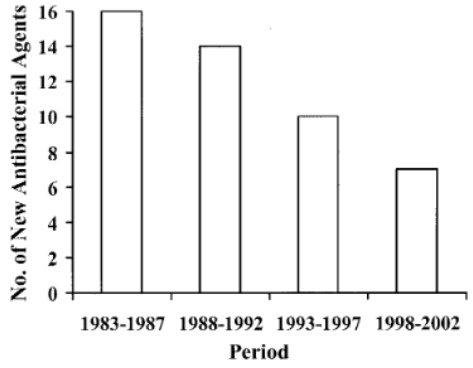

Figure 1. New antibacterial agents approved in the United States, 1983–2002, per 5-year period.

(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.

The bigger picture

Why understand innovation at a drug-pathogen level?

Research statement

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

Notice 2 challenges:

1. The attribution challenge: Attributing patents to the antibiotic class
2. Finding the right comparison group to antibiotic patents

If the drug molecules and pathogens inside each patent is known we can use that to solve both these challenges.

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

With this, x-axis of the plot is done. Note y-axis still remains elusive. A different policy relevant y-axis maybe needed.

Plots proposed

1. Mirroring the burden plot from Murray et al. (2022) as close as possible
   - DATA: drug = *cat(*`Drugs`*)*
     DATA: patho = *cat(*`Pathogens`*)*
     TRANS: inno = *summary.count(*`Patent trends`*)*
     ELEMENT: *polygon(position(bin.rect(*drug*patho*))*, color.hue(*inno*)*, *label(*mort*))*
2. Supplementing with location data of assignees (pharma firms) and inventors (scientists)
   - For this US, this can be done at a state-level with data from the USPTO

... and many more as time permits

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.