Visualizing Innovation in Antibiotics

Project Proposal for Data Visualization (EPPS 6356)

Venkatesh Subramanian*

October 14, 2025

1 Topic: Innovation in Antibiotics

I propose a project to visualize the innovation policy problem underlying Bacterial Antimicrobial Resistance (hitherto, AMR).

The AMR problem has been widely recognized as a global public health "threat" (Centers for Disease Control and Prevention (U.S.) 2019) with 1.27 million deaths worldwide in the year 2019 being attributed to it (Murray et al. 2022). These deaths are caused by the fact that bacteria have evolved to combat against currently available antibiotic drugs while innovation in antibiotics remains insufficient to invent new antibiotics to keep up with the pace of bacterial evolution (Spellberg et al. 2004; Freire-Moran et al. 2011; Zhang et al. 2016; Clerk 2015). Several policy incentives are currently being explored to 'pull' and/or 'push' antibiotic drug innovation further (McDonnell et al. 2023; Renwick, Brogan, and Mossialos 2016) with attempts being made to estimate the appropriate size of such incentives (Towse and Bonnifield 2022; Outterson 2021). To aid in these attempts, I propose to make granular, reproducible visualizations using public available data sources at the level of individual molecular structures found in worldwide patents with potential/known antibiotic action.

*PhD student in Public Policy and Political Economy, School of Economic, Political, & Policy Sciences, The University of Texas at Dallas, 800 W. Campbell Road, Richardson, TX 75080, USA. Email: venkatesh.subramanian@utdallas.edu. Webpage: https://epps.utdallas.edu/phd-students/venkatesh-subramanian/

There are a couple of well-cited examples of existing dashboards that aim to provide evidence for antibiotic innovation. First is the dashboard by Global AMR R&D Hub (2025). Their visualizations do not document patents, and don't provide a visual comparison of innovation across different molecules (which is necessary for my research statement below). Second is the dashboard by Susan Hawes et al. (2025) which explicitly documents patents. But it uses the CPC system used by patent examiners instead of the molecular structures mentioned in the patents as I propose to do here. This can result in selection biases since firms can have strategic reasons not to reveal the therapeutic activity of a new molecule to the patent examiner given that many drugs can be patented before finishing the clinical trial process.

2 Research statement

While this project is aimed to provide descriptive evidence on patented molecules with potential antibiotic action, it brings me one step closer to address a bigger research question: *How much worser are the product market outputs of antibiotic patents than comparable patents with similar inputs?* Notice that an answer to this question would clearly demonstrate the extent of the innovation insufficiency in antibiotics, and this can useful to reckon the size of required incentives.

Methodologically, this research question faces two challenges. First is the attribution challenge of classifying patents attributable to the antibiotic class. Second is finding non-antibiotic patents similar to antibiotic patents (i.e) the challenge of identifying the right comparison group. Notice that both problems can be resolved by identifying the molecules in the patent. Once the molecules in the patent are known, I can use it to classify patents as belonging to the antibiotic class. Then, using well-defined measures of molecular similarity, I can find patents with molecules similar to antibiotic molecules.

Yet another challenge here is ensuring comparisons only among patents with 'similar inputs'. This is relegated for a future research project.

Figure 1: Visualization of AMR burden taken from Murray et al. (2022)

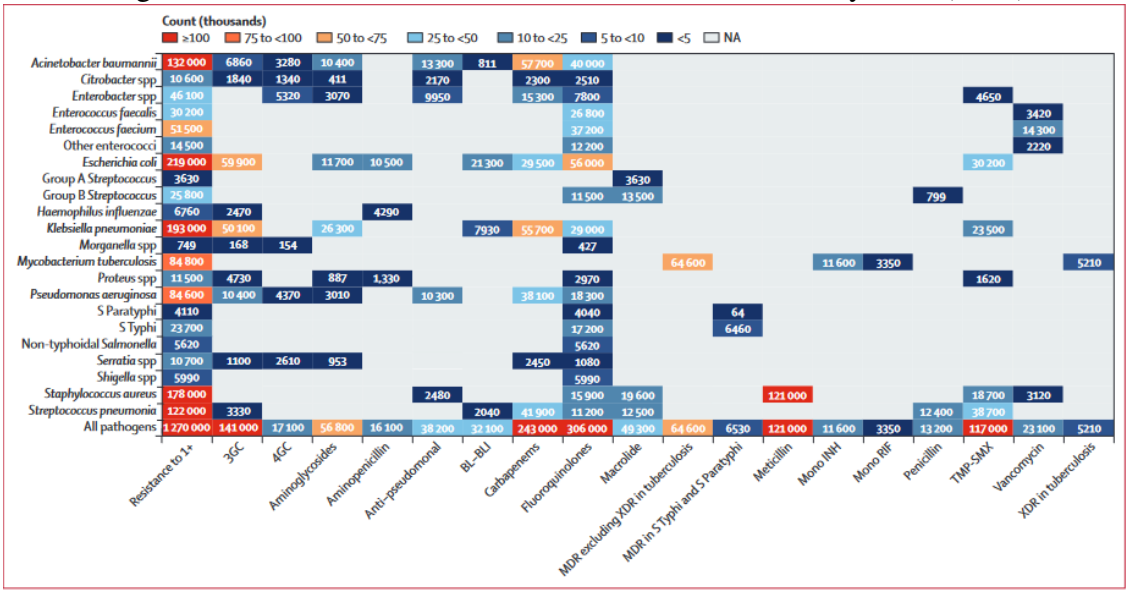

Figure 6: **Global deaths (counts) attributable to bacterial antimicrobial resistance by pathogen–drug combination, 2019**
For this figure, only deaths attributable to resistance, not deaths associated with resistance, are shown due to the very high levels of correlation for resistance patterns between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

3 Method

The proposal here can be brought to fruition using two data sources both of which are publicly available to download. First is SureChEMBL (Papadatos et al. 2016), a cheminformatics database that uses machine learning techniques to annotate the molecules mentioned in different sections (abstract, description, claims, attachments) of worldwide patents. Secondly, I use AntibioticDB (Farrell et al. 2018) to identify molecules with potential antibiotic action. AntibioticDB also has self-reported data on which stage of the clinical trial pipeline each molecule was last observed. Both SureChEMBL and AntibioticDB use the International Chemical Identifier (InChi)) key to identify individual molecules, helping me join them both.

I propose to use these two datasets to make a visualization comparable to the visualization of AMR burdens provided in Murray et al. (2022). This visualization is shown in Figure 1. Specifically, I hope to make a matrix visualization that maintains the same x axis as Figure 1. But instead of the numbers/colors describing the mortality, I want to describe innovation proxied by patents. I want my figure to look as similar as possible to Figure 1 so that I can unambiguously motivate why it is significant for policymakers to understand the innovation problem in a granular way. That is, I

want policymakers to be able to see if innovation is worser in those drug molecules that are known to cause higher AMR mortality.

An obstacle that I won't be able to overcome in this project is plotting the data using the same y-axis as Figure 1. This is because SureChEMBL doesn't annotate pathogens mentioned in patents. But even if such data were available, innovation in antibiotics can't be presented with the same grammar. Given that many antibiotic molecules can be broad-spectrum (i.e) can have action against multiple species of bacteria it will be difficult to 'fractionalize' innovation across all of them. This will require me to identify other policy-relevant y-axis.

In addition to the matrix visualization I have described, I hope to make visualizations describing the patent assignees and inventors (i.e) the firms that own the molecules and the researchers who invented them. Given the global coverage of patents in SureChEMBL it should also be possible to make chloropeths describing the spatial variation in patents across countries. If time permits, I can also use data on patents specifically in the US, a major hub of pharmaceutical innovation, sourced from the US Patents and Trademarks Office (USPTO) to understand spatial locaiton of inventors across US states.

Given that this project is aimed at enabling policymakers to make better estimates of the size of policy incentives, reproducibility of analyses will be a cornerstone of this project. With this in mind, throughout this project, I propose to use the ggplot2 R package (Wickham 2016), which uses the Grammar of Graphics framework of Leland Wilkinson (2005). I propose to deploy my visualization using the Shiny framework (Chang et al. 2025). Notice that all the tools here are open-source. In addition, I hope to have my code in a GitHub repository. All this, in conjunction with the public nature of the two datasets, should make this project entirely reproducible by external stakeholders.

References

Centers for Disease Control and Prevention (U.S.) 2019. *Antibiotic resistance threats in the United States, 2019* [in en]. Technical report. Atlanta, Georgia: National Center for Emerging Zoonotic and Infectious Diseases (U.S.), November. Accessed July 10, 2025. https://doi.org/10.15620/cdc:82532. https://stacks.cdc.gov/view/cdc/82532.

Chang, Winston, Joe Cheng, J. J. Allaire, Carson Sievert, Barret Schloerke, Garrick Aden-Buie, Yihui Xie, et al. 2025. *shiny: Web Application Framework for R.* https://github.com/rstudio/shiny.

Clerk, Marks &. 2015. *From rare to routine: Life Sciences Report 2015 on medicines for rare diseases, vaccines and antibiotics.* Technical report. Marks & Clerk. Accessed July 24, 2024. https://marksclerk.b-cdn.net/yrkdn0ye/m-c_lifesciencesreport_2015.pdf.

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.

Freire-Moran, Laura, Bo Aronsson, Chris Manz, Inge C. Gyssens, Anthony D. So, Dominique L. Monnet, and Otto Cars. 2011. "Critical shortage of new antibiotics in development against multidrug-resistant bacteria—Time to react is now." *Drug Resistance Updates* 14, no. 2 (April): 118–124. ISSN: 1368-7646, accessed July 16, 2024. https://doi.org/10.1016/j.drup.2011.02.003. https://www.sciencedirect.com/science/article/pii/S1368764611000197.

Global AMR R&D Hub. 2025. *Global AMR R&D Hub Dashboard,* October. Accessed October 14, 2025. https://dashboard.globalamrhub.org/.

Leland Wilkinson. 2005. *The Grammar of Graphics* [in en]. Statistics and Computing. New York: Springer-Verlag. ISBN: 978-0-387-24544-7, accessed October 15, 2025. https://doi.org/10.1007/0-387-28695-0. http://link.springer.com/10.1007/0-387-28695-0.

McDonnell, Anthony, Katherine Klemperer, Morgan Pincombe, Rachel Silverman Bonnifield, Prashant Yadav, and Javier Guzman. 2023. *A New Grand Bargain to Improve the Antimicrobial Market for Human Health* [in en]. Working Group Report. Center for Global Development. Accessed November 24, 2023. https://pubs.cgdev.org/amr/.

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.

Outterson, Kevin. 2021. "Estimating The Appropriate Size Of Global Pull Incentives For Antibacterial Medicines." Publisher: Health Affairs, *Health Affairs* 40, no. 11 (November): 1758–1765. ISSN: 0278-2715, accessed March 2, 2025. https://doi.org/10.1377/hlthaff.2021.00688. https://www.healthaffairs.org/doi/10.1377/hlthaff.2021.00688.

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.

Renwick, Matthew J., David M. Brogan, and Elias Mossialos. 2016. "A systematic review and critical assessment of incentive strategies for discovery and development of novel antibiotics" [in en]. Number: 2 Publisher: Nature Publishing Group, *The Journal of Antibiotics* 69, no. 2 (February): 73–88. ISSN: 1881-1469, accessed November 19, 2023. https://doi.org/10.1038/ja.2015.98. https://www.nature.com/articles/ja201598.

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.

Susan Hawes, Aaron Ballagh, Mark Garlighouse, Chidananda B G, Chris Angwin, and Ken Williams. 2025. *Antimicrobial Resistance Research and Innovation in Australia* [in en]. Technical report. The Lens, June. Accessed October 14, 2025. https://www.lens.org/lens.

Towse, Adrian, and Rachel Silverman Bonnifield. 2022. *USG Advanced Commitment for Subscription-Based Purchasing of Novel Antimicrobials* [in en]. Policy Paper 277. Center for Global Development, November. https://www.cgdev.org/sites/default/files/ambitious-usg-advanced-commitment-subscription-based-purchasing-novel-antimicrobials.pdf.

Wickham, Hadley. 2016. *ggplot2*. Use R! Cham: Springer International Publishing. ISBN: 978-3-319-24275-0 978-3-319-24277-4, accessed October 15, 2025. https://doi.org/10.1007/978-3-319-24277-4. http://link.springer.com/10.1007/978-3-319-24277-4.

Zhang, Meng, Xiangjun Kong, Jun Zheng, Jianbo Wan, Yitao Wang, Yuanjia Hu, and Rong Shao. 2016. "Research and development of antibiotics: Insights from patents and citation network." *Expert opinion on therapeutic patents* 26 (March). https://doi.org/10.1517/13543776.2016.1167877.