

---

## Application Notes

# A dashboard for exploring clinical trials sponsorship and potential virtual monopolies

S. Scott Graham<sup>1</sup>, Zoltan P. Majdik<sup>2</sup>, Joshua B. Barbour <sup>3</sup>, and Justin F. Rousseau <sup>4,5</sup>

<sup>1</sup>Department of Rhetoric & Writing, Center for Health Communication, University of Texas at Austin, Austin, Texas, USA, <sup>2</sup>Department of Communication, North Dakota State University, Fargo, North Dakota, USA, <sup>3</sup>Department of Communication Studies, Moody College of Communication, University of Texas at Austin, Austin, Texas, USA, <sup>4</sup>Department of Population Health, Dell Medical School, The University of Texas at Austin, Austin, Texas, USA, and <sup>5</sup>Department of Neurology, Dell Medical School, The University of Texas at Austin, Austin, Texas, USA

Corresponding Author: S. Scott Graham, PhD, Department of Rhetoric & Writing, Center for Health Communication, University of Texas at Austin, 208 W. 21st St, STOP B5500, Parlin Hall, Room 3, Austin, TX 78712, USA; ssg@utexas.edu

Received 19 January 2021; Revised 20 May 2021; Editorial Decision 4 October 2021; Accepted 8 October 2021

### ABSTRACT

**Objective:** To create a data visualization dashboard to advance research related to clinical trials sponsorship and monopolistic practices in the pharmaceuticals industry.

**Materials and Methods:** This R Shiny application aggregates data from ClinicalTrials.gov resulting from user's queries by terms. Returned data are visualized through an interactive dashboard.

**Results:** The Clinical Trials Sponsorship Network Dashboard (CTSND) uses force-directed network mapping algorithms to visualize clinical trials sponsorship data. Interpretation of network visualization is further supported with data on sponsor classes, sponsorship timelines, evaluated products, and target conditions. The source code for the CTSND is available at <https://github.com/sscottgraham/ConflictMetrics>.

**Discussion:** Monopolistic practices have been identified as a likely contributor to high drug prices in the United States. CTSND data and visualizations support the analysis of clinical trials sponsorship networks and may aid in identifying current and emerging monopolistic practices.

**Conclusions:** CTSND data can support more robust deliberation about an understudied area of drug pricing.

**Key words:** health policy informatics, data visualization, clinical trials, research funding, network modeling

---

### INTRODUCTION

Americans pay more for prescription drugs than people living anywhere else in the world, and drug pricing has received considerable scrutiny in recent years. Defenders of high prices and price disparities cite research and development (R&D) and marketing expenditures as reasons; however, research indicates that monopolies and monopolistic practices are, in fact, the leading driver of drug prices in the United States.<sup>1–4</sup> For example, pharmaceuticals spending on R&D and marketing is comparable to other industries, but it retains

3.1–12% greater profit margins when compared with similar industries that do not enjoy monopoly protections.<sup>3</sup> Monopolies take multiple forms, including regulated, temporary monopolies that provide patent protection for new and orphan drug products.<sup>1,2,4</sup> Unregulated monopolies eliminate emerging competition and contribute to high prescription drug prices as is the case in industry intervention in the generics market or so-called “capture-and-kill” acquisitions.<sup>5–8</sup> Virtual monopolies, wherein a single or handful of companies control all or nearly all products available for a given

### LAY SUMMARY

The high costs of prescription drugs can have negative effects for patients and society. High prices can prevent some people from being able to access needed care and can be a burden for public healthcare budgets. Recent research indicates that high drug prices are often the result of monopolistic business practices. We developed an interactive RShiny tool that can help researchers and policymakers explore the relationships between pharmaceuticals research funding, drug prices, and industry business practices. The tool allows analysts to create multiple visualizations of clinical trials funding networks drawn from information registered with ClinicalTrials.gov, the US government's official registry of clinical trials. The application can support hypothesis generation and future research into appropriate policy solutions for the high costs of prescription drugs.

condition, have also been linked to drug pricing, most notably in the case of insulin.<sup>1,9</sup> Monopoly disruptions from either so-called “me-too” drugs or the entry of generics into the market reliably reduce prices.<sup>2,3,7</sup>

Concern over the high cost of drugs in the United States has resulted in policy proposals related to reforming patent protections, particularly in the context of the orphan drugs program, and resisting industry efforts to impede the generics market. These proposals address issues identified with regulated and unregulated monopolies while tending to ignore virtual monopolies. If health policy researchers are to understand the effects of virtual monopolies on drug pricing, they need research tools that support efforts to identify them.

Data-driven dashboard solutions have been identified as an important part of health policy research and communication.<sup>10–12</sup> Indeed, the role of such dashboards in health policy decision-making has been especially prominent during the COVID-19 pandemic.<sup>13,14</sup> The ConflictMetrics<sup>15</sup> team developed the Clinical Trials Sponsorship Network Dashboard (CTSND) to understand and visualize the complex funding networks supporting clinical trials and to advance future research in this area. Once identified, subsequent research can evaluate if virtual monopolies are associated with higher prices and assess if virtual monopolies are associated with other negative externalities, such as disparities in access to care and differential efficacy profiles.

## MATERIALS AND METHODS

The CTSND is designed to make clinical trials sponsorship and related health policy research more practical and accessible. By leveraging ClinicalTrials.gov, the CTSND allows users to visualize trial sponsorship, evaluated products, and target conditions data. US federal law and related regulations require registration of clinical trials conducted in the United States and/or related to any FDA-regulated products (42 CFR Part 11). ClinicalTrials.gov is an online service from the National Library of Medicine (NLM) that provides access to registry and clinical trial results data. The site indexes data on interventions assessed, target conditions, study design, trial sponsorship, and, in some cases, research results including adverse events.<sup>16</sup> ClinicalTrials.gov is most known for its web interface, but the NLM also supports bulk data downloads and an application programming interface (API). The API allows free-form query submission and returns results according to the site's relevance ranking algorithm. Results can be retrieved in a variety of data formats including XML, JSON, CSV, and tree. The CTSND uses these resources to make ClinicalTrials.gov sponsorship data more accessible to those who do not have advanced data querying skills. User-generated queries are submitted via the ClinicalTrials.gov API. The API returns a JSON file with data on relevant clinical trials. The

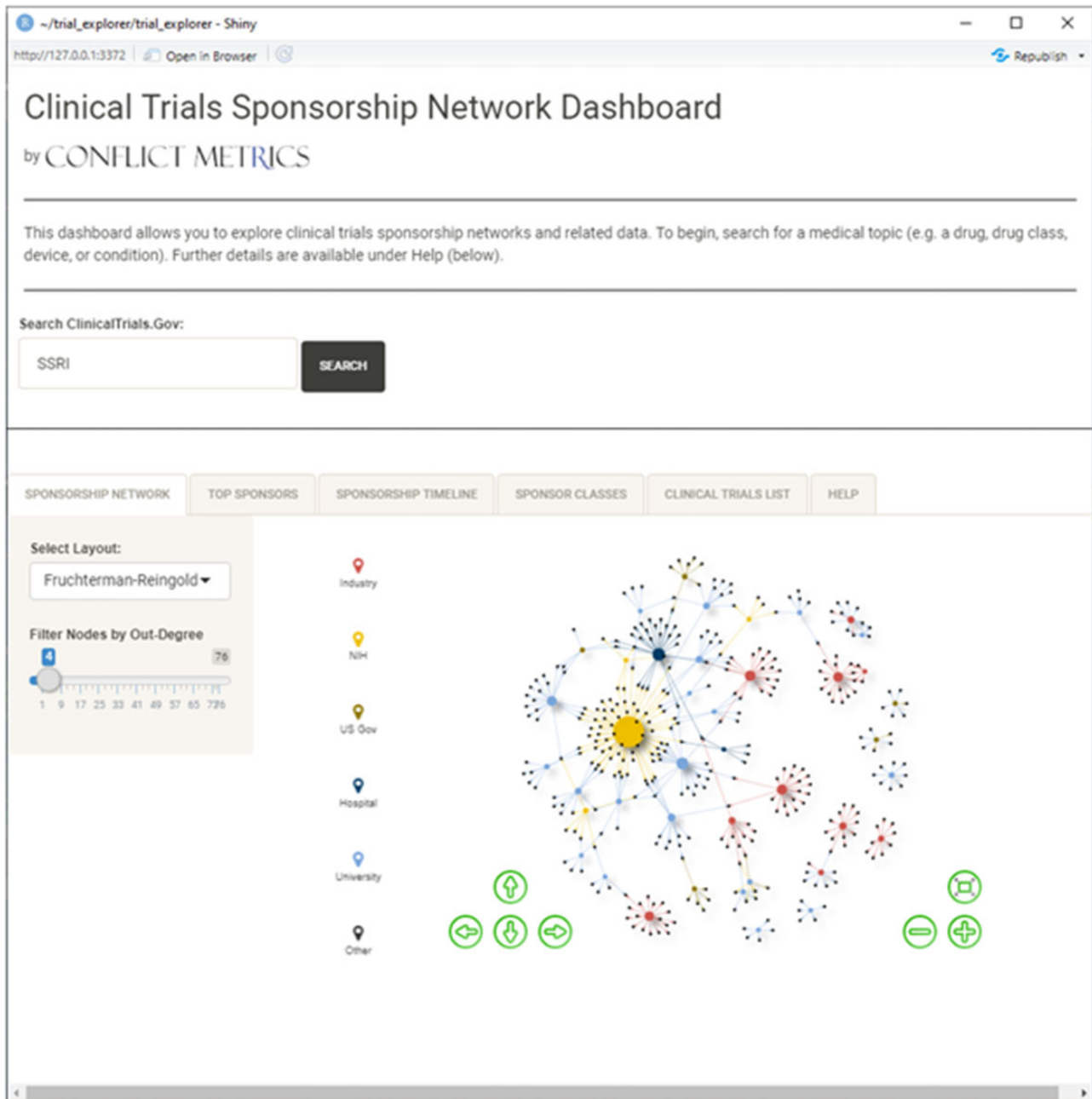
CTSND query pulls the unique trial ID number, brief title, evaluated conditions, evaluated interventions, trial start date, trial completion date, and trial sponsorship data for the 1000 most relevant trials matching the query.

The CTSND is an R Shiny application<sup>17</sup> developed by repurposing a pre-existing conflicts of interest visualization dashboard that had been developed by ConflictMetrics with support from the National Endowment for the Humanities and the National Science Foundation's Extreme Science and Engineering Discovery Environment (XSEDE).<sup>18</sup> CTSND displays data on 5 tabs: Sponsorship Network, Sponsor Classes, Sponsorship Timeline, Top Sponsors, and Trials List. The CTSND generates data visualizations using visNetwork,<sup>19</sup> ggplot2,<sup>20</sup> and streamgraph<sup>21</sup> packages. The Sponsorship Network tab is the signature display of the CTSND (see Figure 1). The tab displays a network map that visualizes the relationships among unique sponsors and individual clinical trials. By default, the dashboard displays network diagrams using the Fruchterman–Reingold layout algorithm, a force-directed graph algorithm that simulates spring-like attractive forces between network nodes, minimizing the overlap of nodes and edges.<sup>22</sup>

## RESULTS

Data visualization is a key tool in the suite of techniques that should be employed during data gathering, analysis, and research. The results illustrate the usefulness of the tool by highlighting what analysts should consider in their interpretative work through examples of already well-studied monopolistic practices. Analysts should look for patterns that suggest the concentration of influence in a single or just a few entities (ie, a dominant node or set of nodes), which could be indicative of monopolistic practices and inspire further research.

Force-directed algorithms simulate influence as a physical property (node size) and are thus especially useful when trying to identify dominant nodes or node clusters within a network of sponsors and products. For example, a small number of larger and thus more influential sponsor nodes may indicate the presence of regulated, unregulated, or virtual monopolies. Analysts can also select from among several common layout algorithms and use the out-degree filter to reshape the diagram as needed. “Out-degree,”<sup>23</sup> here refers to the number of clinical trials supported by a given sponsor. An out-degree filter of 4, for example, would limit the entities displayed to only include those nodes (sponsors or trials) where any given sponsor has supported 4 or more trials. The dynamic network display allows direct manipulation (drag-and-drop, zoom, pan) so users can explore data. This sort of engagement may encourage more thoughtful, active exploration, because viewers can rearrange the visual as their use inspires them. The Dashboard parses sponsor names and



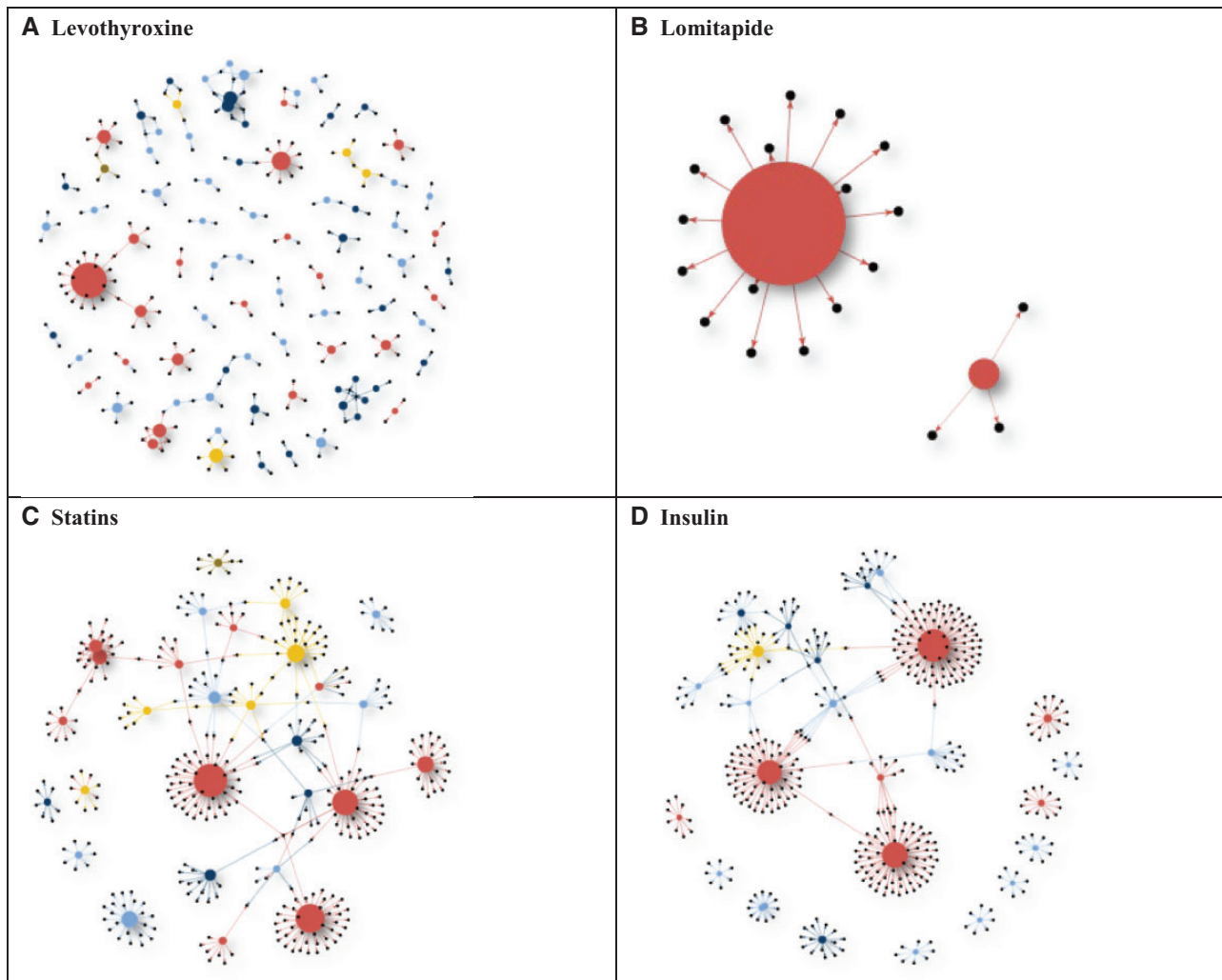
**Figure 1.** Screen capture of CTSND displaying sponsorship network for selective serotonin reuptake inhibitors (SSRIs). The network is displayed using the Fruchterman–Reingold algorithm and the out-degree filter is set to 4.

canonical classes to sort sponsorships into hospital, industry, university, NIH and other US government, and other categories. Sponsorship class data can allow users to identify when research in a given area is dominated by industry or when it is the result of a mix of industry, university, and federal sources. Sponsor nodes are color-coded by these ConflictMetrics.com-defined sponsor classes.

The Top Sponsors tab provides tabular data identifying the name, class, and number of sponsorships in descending order. A “long tail” or the presence of many sponsors with the majority of the trials sponsored by just a few companies, may also be an indication of virtual monopoly in a research area. The Sponsorship Timeline tab offers a streamgraph of sponsorship over time for the top 30% of sponsors (see Figure 3). The streamgraph may be useful for

identifying potential capture-and-kill acquisitions by highlighting bottlenecks in the makeup of funding. The Sponsorship Classes tab display the percentage of sponsors according to ClinicalTrials.gov and ConflictMetrics.Com sponsor classes. The canonical ClinicalTrials.gov sponsor classes include, the federal government, industry, hospital networks, the NIH, other, and other (non-US) governments.

The CTSND also outputs network and supporting data to support analyses of sponsorship networks. These data may also support subsequent analyses of drug pricing based on the prevalence of industry–university and industry–government partnerships. Finally, the Clinical Trials List provides additional contextual information to support analysis of data displays on other tabs. The Clinical Tri-



**Figure 2.** CTSND network visualizations for levothyroxine (A), lomitapide (B), statins (C), and insulin (D). Black nodes are sponsored clinical trials. Sponsoring nodes are color-coded by sponsor type (red for industry, light blue for university or hospital, yellow for U.S. government, and dark blue for others).

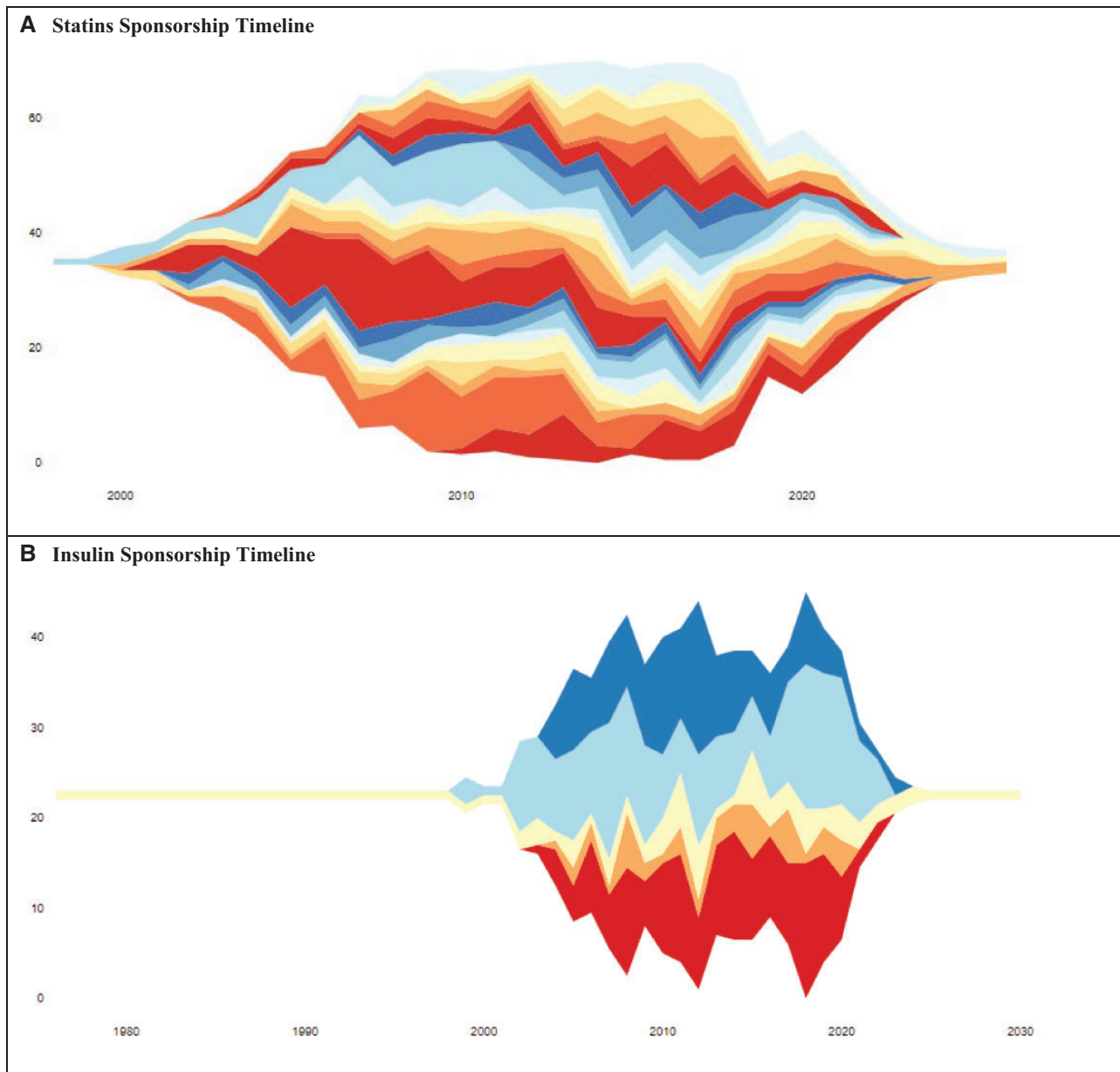
als Data tab identifies the most evaluated interventions and conditions, which is also useful in identifying potential virtual monopolies. If many different products are under active evaluation, then the search query is unlikely to have identified a virtual monopoly. If just a few products are being assessed by a small number of companies, then the search may have identified a virtual monopoly.

## DISCUSSION

A few illustrations demonstrate how these data can support the identification of potential virtual monopolies. Figure 2 compares network visualizations for several targeted searches. Each of these visualizations were selected because they illustrate certain network signatures that may be associated with different market structures. Levothyroxine clinical trials sponsorship (Figure 2A) is characterized by a diffuse, network with many discrete clusters of nodes. Levothyroxine is a commonly used prescription drug products manufactured by a number of companies. Levothyroxine sponsorship is correspondingly diffuse, with multiple trials sponsored by companies and federal agencies, hospitals, and universities. In contrast, the lomitapide (Juxtapid) sponsorship network (Figure 2B) is

characterized by a relatively small number of trials and 1 key sponsor. The single, large node indicates a sponsorship monopoly, and lomitapide is one of the most expensive drugs available and benefits from regulated monopoly status until 2032.<sup>24</sup> All trials are sponsored by Aegerion Pharmaceuticals, Inc. (large node) and Amryt Pharma (small node) who acquired Aegerion in 2019.

While the generics and regulated monopoly markets are generally well-understood in terms of pricing effects, the CTSND allows for more detailed explorations of alternative market configurations. For example, the statins sponsorship network (Figure 2C) shows a hybrid profile with integrated and diffuse regions showing multiple significant industry sponsors, each sponsoring a number of trials. Statins are a widely cited example of “follow-on” or “me-too” drugs. Me-too drugs are competitor reformulations that can be brought relatively quickly to market following the introduction of a successful new drug class. Competitive pressures brought by me-too drugs have been shown to result in significant cost savings for payors.<sup>25</sup> This is not always the case, however. For example, insulin—often identified as a virtual monopoly—has a sponsorship network (Figure 2D) with a similar profile to the statin network—hybrid with diffuse and integrated sections showing several key industry sponsors. But insulin’s



**Figure 3.** Sponsorship timelines for statins (A) and insulin (B). Each unique sponsor is represented by its own line, and each line is assigned 1 of 6 distinctive colors which are automatically arrayed to maximize an analyst’s ability to visually distinguish among sponsors. The vertical thickness of the line ( $y$ -axis) represents the number of trials supported by the sponsors over time. The scale of the  $y$ -axis is set by the moment in time with most sponsored trials. In the interactive application, hovering over each band shows the name of the sponsor and the number of trials at any given moment on the  $x$ -axis.

network displays with a higher degree of concentration among 3 sponsors (Novo Nordisk, Ely Lilly, and Sanofi). The key network signature differences are the total number of dominant industry sponsors (6 vs 3) and the distribution of trials per sponsor.

These visuals should inspire and guide questions and further research. For example, analysts working with these visuals could ask if the differences between Figure 2C and D are meaningful compared to the striking dominance of sponsors in Figure 2B. Interpreting multi-dimensional data represented in 2 dimensions is difficult, and CTSND helps by allowing for interactive reconfiguration of the visual in real time. The utility of reconfiguration of the visuals should be judged to the extent that they reveal possible patterns for further research.

The CTSND streamgraph also supports such analyses adding the dimension of time. Figure 3 compares the statin and insulin sponsorship timelines. The near-simultaneous proliferation of trails across multiple sponsors (Figure 3A) is likely a hallmark of drug classes with a rich follow-on market. Insulin sponsorship (Figure 3B), again, features a much smaller number of sponsors conducting a greater proportion of trials per sponsor. Contraction at the right edge of plots is often an artifact of current trial registrations. Most registered trials are scheduled to terminate within 6–7 years of their start date. The goal of these data visualizations is not to allow for drawing final conclusions about sponsorships in a vacuum, but to support the exploration of patterns over time through complementary data analysis and research.



## CONCLUSION

In sum, additional research is needed to address the high costs of prescription drugs in the United States. An essential part of this effort will involve determining the factors most likely to predict high costs. Future work leveraging the CTSND and ClinicalTrials.gov data can provide insights into direct and indirect effects of conflicts of interest on biomedical research and health outcomes. Virtual monopolies are increasingly identified as a significant contributor to high costs, the CTSND can advance the foundational research and policy solutions focused on virtual monopolies.

The CTSND, as a visualization tool, will be most helpful as a part of initial exploratory research and hypothesis generation. The underlying network modeling technology, however, can be leveraged to quantify network properties and their relationships with measures of pricing indices or market share to support health policy informatics research. For example, this tool could support the examination of the relationship between trial sponsorship and market share in the context of COVID-19 therapeutics and vaccines. The CTSND and its underlying network modeling technologies can advance the foundational research and policy solutions in this area by providing dynamic visualization of multidimensional network relationships to support the rigorous computational analyses of these networks.

## FUNDING

This work was supported by the National Endowment for the Humanities (HAA-261070) and the National Science Foundation's Extreme Science and Engineering Discovery Environment allocation number HUM180003.

## AUTHOR CONTRIBUTIONS

SSG and JFR were responsible for initial conceptualization of this article. SSG developed the dashboard application and wrote the first draft of the manuscript. JBB, JFR, and ZPM provided feedback on the manuscript, participated substantively in revision, and approved the final version.

## CONFLICT OF INTEREST STATEMENT

None declared.

## DATA AVAILABILITY

The source code for the CTSND is available at <https://github.com/scottgraham/ConflictMetrics>.

## REFERENCES

- Rajkumar SV. The high cost of prescription drugs: causes and solutions. *Blood Cancer J* 2020; 10 (6): 71.
- Siddiqui M, Rajkumar SV. The high cost of cancer drugs and what we can do about it. *Mayo Clin Proc* 2012; 87 (10): 935–43.
- Patterson JA, Carroll NV. Should the United States government regulate prescription prices? A critical review. *Res Social Adm Pharm* 2020; 16 (5): 717–23.
- Murphy SM, Puwanant A, Griggs RC; Consortium for Clinical Investigations of Neurological Channelopathies (CINCH) and Inherited Neuropathies Consortium (INC) Consortia of the Rare Disease Clinical Research Network. Unintended effects of orphan product designation for rare neurological diseases. *Ann Neurol* 2012; 72 (4): 481–90.
- Cunningham C, Ederer F, Ma S. Killer acquisitions. SSRN 3241707. March 22, 2019. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3241707](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3241707) Accessed October 17, 2021.
- Alpern JD, Stauffer WM, Kesselheim AS. High-cost generic drugs—implications for patients and policymakers. *N Engl J Med* 2014; 371 (20): 1859–62.
- Dave CV, Kesselheim AS, Fox ER, Qiu P, Hartzema A. High generic drug prices and market competition: a retrospective cohort study. *Ann Intern Med* 2017; 167 (3): 145–51.
- Jones GH, Carrier MA, Silver RT, Kantarjian H. Strategies that delay or prevent the timely availability of affordable generic drugs in the United States. *Blood* 2016; 127 (11): 1398–402.
- Rajkumar SV. The high cost of insulin in the United States: an urgent call to action. *Mayo Clin Proc* 2020; 95 (1): 22–8.
- Gourevitch MN, Athens JK, Levine SE, Kleiman N, Thorpe LE. City-level measures of health, health determinants, and equity to foster population health improvement: the City Health Dashboard. *Am J Public Health* 2019; 109 (4): 585–92.
- Foldy SL, Biedrzycki PA, Baker BK, et al. The public health dashboard: a surveillance model for bioterrorism preparedness. *J Public Health Manag Pract* 2004; 10 (3): 234–40.
- Gee RE, Muncy ML. Dashboards to support state health policy making. *JAMA Health Forum* 2020; 1 (4): e200469.
- Arora RK, Joseph A, Van Wyk J, et al. SeroTracker: a global SARS-CoV-2 seroprevalence dashboard. *Lancet Infect Dis* 2021; 21 (4): e75–6.
- Wissel BD, Van Camp PJ, Kouril M, et al. An interactive online dashboard for tracking COVID-19 in U.S. counties, cities, and states in real time. *J Am Med Inform Assoc* 2020; 27 (7): 1121–5.
- Graham SS. About ConflictMetrics. 2020. <http://conflictmetrics.com/about> Accessed October 17, 2021.
- Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database—update and key issues. *N Engl J Med* 2011; 364 (9): 852–60.
- Chang W, Cheng J, Allaire J, Xie Y, McPherson J. Shiny: web application framework for R. R Package Version. 2017; 1 (5). <https://cran.r-project.org/web/packages/shiny/index.html> Accessed October 17, 2021.
- Graham SS, Majdik ZP, Clark D. Methods for extracting relational data from unstructured texts prior to network visualization in humanities research. *J Open Hum Data* 2020; 6 (1): 8.
- Almende BV, Thieurmél B, Robert T. visNetwork: Network visualization using 'vis.js' library. 2019. <https://cran.r-project.org/package=visNetwork> Accessed October 17, 2021.
- Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer; 2016.
- Rudis B. streamgraph. 2018. <https://github.com/hrbrmstr/streamgraph> Accessed October 17, 2021.
- Fruchterman TM, Reingold EM. Graph drawing by force-directed placement. *Softw Pract Exp* 1991; 21 (11): 1129–64.
- Borgatti S, Mehra A, Brass DJ, Labianca G. Social network analysis in the social sciences. *Science* 2009; 323 (5916): 892–5.
- Jiang K. Near-record number of approvals signals drug development shift. *Nat Med* 2013; 19 (2): 114.
- Régnier S. What is the value of 'me-too' drugs? *Health Care Manag Sci* 2013; 16 (4): 300–13.