

The Public Impact of Using Observational Data to Enhance Efficiency and Effectiveness: From COVID-19 Vaccines to Startups to Emergency Rooms

Soroush Saghafian

June 2024

Blog Series: [PUBLIC IMPACT ANALYTICS SCIENCE \(PIAS\)](#)



Photo Credit: DALLE 2023

In February of 2021, a significant problem was on top of mind for many officials across the globe: are recently developed COVID-19 vaccines effective enough to end the terrible tragedy caused by the pandemic.

I was thinking about the same issue during the same time. There was some evidence from randomized control trials, but the evidence was very limited. For example, the trials did not include enough people from different races, background, conditions, and ages, among others. So it was not clear whether the limited results showing that vaccines are (at least to some extent) effective can be extended to the general public.

While thinking about this on February 24 of that year, I found myself browsing some tweets on X (which was called Twitter then). Interestingly enough, I came across a tweet by one of my colleagues at Harvard:

“We've just confirmed the effectiveness of the Pfizer-BioNTech vaccine outside of randomized trials.”

The tweet was from Miguel Hernan, a professor of epidemiology at Harvard T.H. Chan School of Public Health and the director of Causal Lab. He had linked to their recently published paper by the *New England Journal of Medicine* [1]. The paper started by emphasizing this:

“As mass vaccination campaigns against coronavirus disease 2019 (Covid-19) commence worldwide, vaccine effectiveness needs to be assessed for a range of outcomes across diverse populations in a noncontrolled setting.”

Passages like this remind us about the colossal value observational data can have for moving our societies forward. They also remind us about “referring to matters of observation and experiment” when discussing “adequacy of evidence,” as the quote from John Dewy—an American philosopher, psychologist, and educational reform—emphasizes:

“If a scientific man be asked what is truth, he will reply ... that which is accepted upon adequate evidence. And if he be asked for a description of adequacy of evidence, he certainly will refer to matters of observation and experiment.”

Specifically, the study of COVID-19 vaccine effectiveness by Hernan and his co-authors allowed them to use observations coming from a “noncontrolled setting” as opposed to data stemming from an experiment, enabling them to benefit from data on a much larger and more diverse scale.

Beyond vaccines, COVID-19 observational data also happened to be invaluable in finding superior ways of improving the overall efficiency and effectiveness of the healthcare sector. As an example, and as we argue in a recent publication [2], while COVID-19 accelerated the rate of closure of hospitals (some which were arguably not financially and/or operationally efficient), the observational data it provided created unique opportunities for researchers to inform policymakers by conducting careful studies that can shed light on different implications, trade-offs, and consequences of various strategies that can improve the overall efficiency of the healthcare sector.

Putting COVID-9 aside, a more fundamental problem in the healthcare sector is that we do not know which providers are efficient and effective. For example, we do not have good measures or even accepted ways of gauging the effectiveness and efficiency of those who are at the forefront of care delivery. How can one try to improve the efficiency or effectiveness of a super complex system like the U.S. healthcare (which per capita spends about 145% more than the Organization for Economic Cooperation and Development (OECD) median) without even having a good way of measuring the efficiency or effectiveness of its providers?

Using observational data can be very informative in understanding which providers are effective and efficient. And this is what we have done in my lab [3, 4, 5]. A particularly simple though important

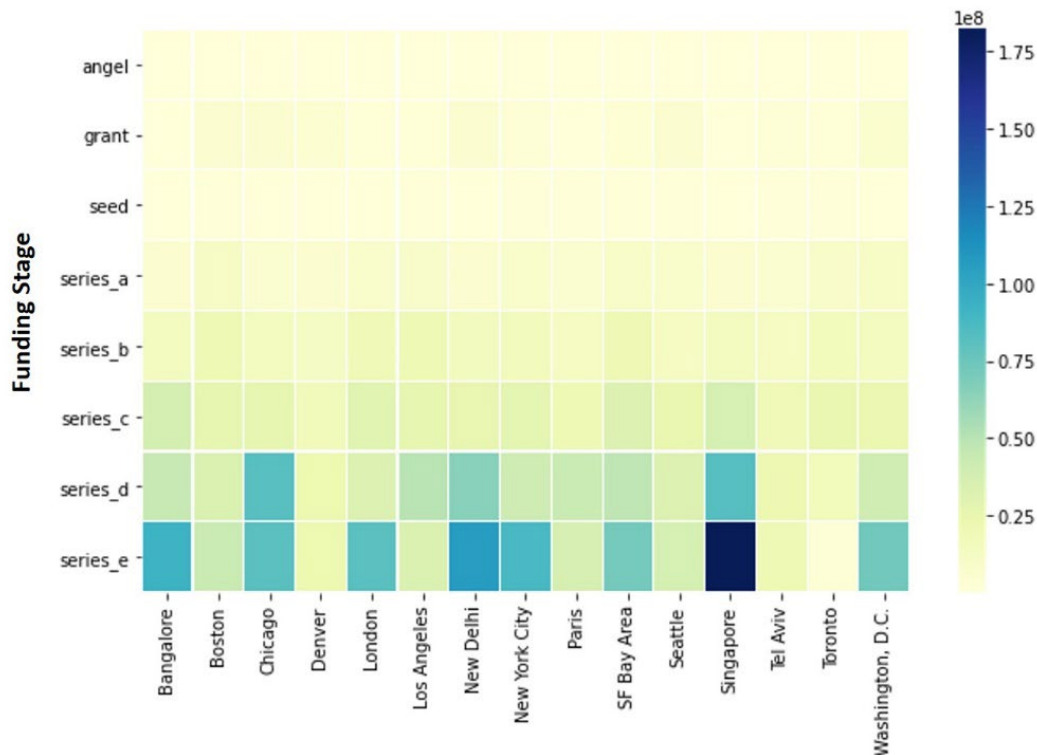


Figure: Heatmap of startup funding (in logarithmic scale) in top 15 cities of the world [Source: our data and analysis in [11]]

question we have been trying to answer is this: who is an effective and efficient physician? How should we train physicians that are not effective or efficient to become an effective and efficient physician.

We have also identified large-scale trends in the healthcare sector that are causing negative impacts. One is the substantial and rapid change with hospitals purchasing many physician practices—a phenomenon often called “vertical integration.” Using large-scale observational data of 2.6 million patient visits across 5,488 physicians, we found various negatively consequences of vertical integration, and also shed light on potential mechanisms through which policymakers can mitigate such consequences [6, 7]. Large-scale observational data is also significantly informative in investigating the inefficient and ineffective use of resources in healthcare caused by the large misalignment between provider capabilities and patient needs, and in designing policies that can address them [8, 9, 10].

But, of course, healthcare is not the only sector in which observational data can have a colossal impact in improving efficiency or effectiveness. We can also see the vital role of observational data in enhancing efficiency in various other domains. Let us consider, for example, the startup ecosystem.

A few years ago, a couple of my students and I decided to think about whether and what we can do to improve this ecosystem. Startups play an important role in the general health of the economy, including technological innovation, economic growth, and jobs creation. They are often thought of as a panacea for solving unemployment and a catalyst for growth. Thus, most cities around the work want to attract startups that can be successful and grow. The investors also want to identify the startups with potential for success, and help them grow.

The problem, however, is that a huge amount of money is spent in startups through various rounds and sources of funding across the globe, but only a tiny fraction of the startups end up being successful. Success stories are indeed rare events in the startup world, with about 90% of startups failing on average. Similarly, as high as 75% of venture-backed deals typically fail to return the investment. In short, a lot of money spent by investors, policymakers, local and state governments, among others, goes to startups that end up failing. For someone like me who has spent most of his entire career worrying about inefficiencies and creating a better world by removing them, this seemed unacceptable. The world, at least in my view, would be a much more efficient place, if we could somehow figure out in advance which startups will be successful and avoid wasting resources in others. So a couple of my students and I started to do some research to find out what can be done.

As part of our study, we were able to obtain detailed data of the amounts and types of investments made, including over 29k startups funding instances around the world. The figure in the previous page is based on our work, and shows the heatmap of startup funding (in logarithmic scale) in the top 15 cities of the world that have the highest number of startups. The darker colors are instances with higher funding amounts raised. The early-stage rounds are approximately similar across the cities in this figure, except for New Delhi that has more funding in terms of grants (very early funding stage). Boston and San Francisco have more Series A funding, while Chicago, London, Singapore, and Bangalore have higher Series E (late stage) funding compared to other cities.

A natural question we had was this: can we develop an algorithm and train it on our data to reliably predict, based on early-stage characteristics of a startup, whether it will end up being successful? We found—using Deep Learning—that the answer is yes. We observed, for instance, an accuracy of over 92% on test data sets. What is more, we were able to generate important insights into what can be done by entrepreneurs, investors, and policymakers or government officials seeking to improve the startup ecosystem and remove various existing inefficiencies. Our recently published book chapter “Using Machine Learning to Demystify Startups’ Funding, Post-Money Valuation, and Success” provides recommendations in this regard [11].

With the ever-growing amount of observational data, it is upon us to find ways in which we can address the important societal challenges of our era without waiting for ideal experimental data to magically show up. Wasting societal resources through inefficient and ineffective mechanisms, and blaming it on the lack of experimental data is an inexcusable mistake. Of course, using observational data without enough care can yield wrong insights. But we do have the means to understand what can go wrong and avoid them.

References

- [1] Dagan, N., Barda, N., Kepten, E., Miron, O., Perchik, S., Katz, M. A., ... & Balicer, R. D. (2021). BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting. *New England Journal of Medicine*, 384(15), 1412-1423.
- [2] Saghafian, S., Song, L. D., & Raja, A. S. (2022). Towards a more efficient healthcare system: Opportunities and challenges caused by hospital closures amid the COVID-19 pandemic. *Health Care Management Science*, 25(2), 187-190.

- [3] Saghafian, S., Imanirad, R., & Traub, S. (2018). Who is an Efficient and Effective Physician? Evidence from Emergency Medicine. *Working Paper, Harvard University*. Available at SSRN.
- [4] Saghafian, S., Imanirad, R., & Traub, S. J. (2019). Do Physicians Influence Each Other's Performance? Evidence from the Emergency Department. *Working Paper, Harvard University*. Available at SSRN.
- [5] Jameson, J., Saghafian, S., Huckman, R. S., & Hudgson, N. (2024). Variation in Batch Ordering of Imaging Tests in the Emergency Department and the Impact on Care Delivery. *Working Paper, Harvard University*. Available at SSRN.
- [6] Saghafian, S., Song, L., Newhouse, J., Landrum, M. B., & Hsu, J. (2023). The impact of vertical integration on physician behavior and healthcare delivery: Evidence from gastroenterology practices. *Management Science*, 69(12), 7158-7179.
- [7] Shah, E. D. (2023). Commentary on "The Impact of Vertical Integration on Physician Behavior and Healthcare Delivery: Evidence from Gastroenterology Practices". *Management Science*, 69(12), 7180-7181.
- [8] Saghafian, S., & Hopp, W. J. (2020). Can public reporting cure healthcare? the role of quality transparency in improving patient-provider alignment. *Operations Research*, 68(1), 71-92.
- [9] Saghafian, S., & Hopp, W. J. (2019). The role of quality transparency in health care: Challenges and potential solutions. *National Academy of Medicine (NAM)*, perspectives.
- [10] Atkinson, M. K., & Saghafian, S. (2023). Who should see the patient? On deviations from preferred patient-provider assignments in hospitals. *Health Care Management Science*, 26(2), 165-199.
- [11] Ang, Y. Q., Chia, A., & Saghafian, S. (2022). Using machine learning to demystify startups' funding, post-money valuation, and success (pp. 271-296). *Springer International Publishing*.