VIEW POINT

HOW BIG DATA ANALYTICS CAN PREVENT AND MANAGE FUTURE PANDEMICS?

Infosys®
Navigate your next

“We are not ready for the next epidemic!” These were the words of Bill Gates at a TED conference back in 2015, where, he took the example of Ebola outbreak and outlined the reasons why we were not ready to fight against the next epidemic.[1]

In his address, he also mentioned the reasons why Ebola did not spread much. Firstly, because of a lot of heroic work by the frontline workers; secondly because of the nature of the virus; it does not spread via air and by the time the patient gets contagious, they are already bedridden. And the third reason was, it did not get into the urban areas; which was just pure luck.[1] In 2020, we are not as lucky like he predicted! People who were affected by the COVID 19 infection, were already contagious, travelled globally and here we are, fighting against a pandemic, which has affected about 32.3 Million people all over the world and has caused about 983000 deaths so far![6] If all stakeholders had taken hint from his address and built a global alert and response system, we would have been in a better shape to fight against COVID-19. Unfortunately, we are still where we were five years back. We are without strong healthcare systems to respond quickly to an outbreak. We don’t have medical core reserves paired with military to respond quickly, move fast and secure areas. We don’t carry out simulations to face such situations. These are some of the key pieces of a global alert and response system that Bill Gates talked about in his address.[1]

But, what do we have now? We have the science and technology at our disposal which we did not have back during the Spanish flu in 1918, which then affected about one third of the world’s population.[1][9] We had to learn it the hard way, but, with appropriate use of the resources and knowledge we have now, we can prevent future pandemics. One of the main tools to fight against future pandemics is – Big Data Analytics. When we look at data as a fighting tool, not only the data generated by a data intensive industry like the Pharmaceutical industry is going to be alone useful, but, data generated by hospitals, research institutes, medical device companies, government agencies, world agencies, also patients and individuals themselves is to be looked at to derive insights and take timely actions to prevent pandemics. Structured, unstructured, traditional and nontraditional, all types of data can help create alerts in the systems, if used judiciously.[2]

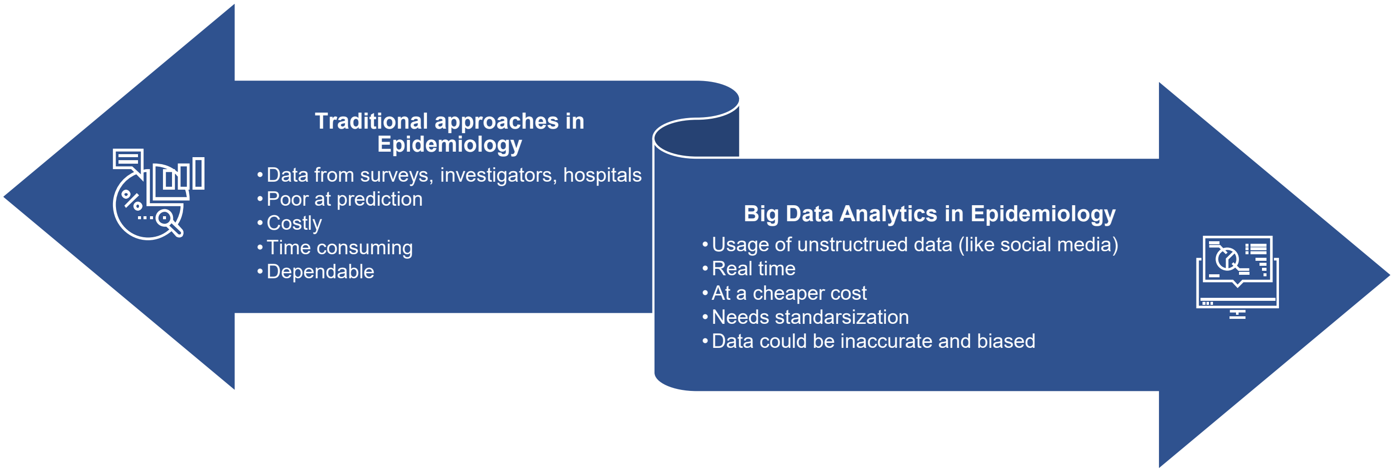
One of the key pieces of a global alert and response system is ramped up research and development.[1] Currently, the areas where researchers are seeing potential of using big data analytics in the pharmaceutical industry are [3]:

- Accelerating drug discovery and development process
- Optimizing the process of clinical trials with respect to cost and speed; analyzing remote patient monitoring data, examining past clinical trials data
- Analyzing genomic data to create personalized medicines
- Making drug cost decisions
- Sifting data from consumers to understand consumer preferences and needs
- Creating effective sales and marketing strategies
- Creating timely alerts to reduce compliance failures
- Optimizing manufacturing processes through the analysis of data points to predict risks such as quality issues; managing demand supply

When we look at using Big Data Analytics in the pharma industry, the key area of its application is in accelerating the process of drug discovery and development in case of an outbreak. Not only we need breakthroughs in vaccines which can cover a range of microbes, but we also need a robust system which enables quick development of vaccines and medicines when an outbreak is announced. Promising drug candidates could be identified from the huge amount of data that is available using Big Data Analytics. Data sharing among companies during an outbreak is an essential aspect to accelerate this process. Currently, companies are working in the domain of cancer research to share past data on a common platform and such collaborations are required during an epidemic too.[3] This lack in robust data sharing platforms in health care has become more obvious while fighting against COVID-19. This is because, not only data sharing among companies is required to rapidly develop medicines, but we also need platforms for pooled, publicly available and deidentified patient level data to draw inferences to update epidemiological investigations, guide treatment protocols and respond to a rapidly evolving situation when data from clinical trials is generating slowly or does not exist.[4] This lack was realized worldwide and efforts are being made by various agencies to build such datasets to make headway in the current situation as well as be ready for the future. For e.g. National COVID Cohort Collaboration has been launched to create secure and centralized repository of data from medical records of COVID – 19 patients across the United Stated of America. This will be used for big data studies of the disease. The goals of this effort are to harmonize the medical records in common data models, identify interventions to fight against the pandemic, understand the long-term impact of the pandemic and demonstrate that collaborative approach shall help in such situations in the future.[12]

Technology companies like Amazon Web Services, Google Cloud and Verizon Media are currently encouraging and facilitating the use of Big Data Analytics to fight against COVID – 19. Amazon Web Services is providing free access in Amazon S3 to AWS COVID -19 data lake. The data lake has up-to-date, publicly readable, curated datasets mostly sourced through AWS data exchange. The data can be used to develop solutions, can be used at organizational level to deploy resources or at societal level to forecast hotspots and trends. Datasets available in the data lake are The New York Times (case tracking data); COVID Tracking Project (testing data); Definitive Healthcare (hospital bed availability data); Delphi Research Group (health survey); Allen Institute for AI (data from more than 45,000 research articles).[13] The Google Cloud Public Datasets Program hosts a repository of public datasets related to COVID-19; is free for access and analysis for researchers. With their partnership with data providers, they have been able to onboard data from various data sources to BigQuery (a serverless, multi-cloud data warehouse). Datasets on BigQuery are The New York Times (helping estimate true toll of Pandemic, understanding role of mask wearing); COVID-19 Community Mobility Reports (data on movement trends); COVID-19 Public Forecasts (to project case counts and deaths); American Community Survey; OpenStreetMap; U.S. Area Deprivation Index (measure of community vulnerability); American Hospital Association (determines ability to handle surge in hospitalizations); Immune Epitope Database (to investigate immune response); COVID-19 Open Data dataset (combines publicly available datasets).[14] Verizon Media has developed an open-source search engine whereby they have indexed more than 44,000 articles by keywords. CORD-19 - COVID-19 Open Research Dataset is the dataset on which this search engine has been built. Research community can apply text and data mining to derive insights.[15]

Development of such databases with large amounts of aggregated data and such new approaches to leveraging big data can improve disease surveillance and enable a timely response to outbreaks which can help in spotting problems at an early stage and prevent small incidents to become pandemics in the future.[2] This in turn can create an efficient global alert and response system. Preventing the lag in detection is the key to prevent future pandemics.[2]



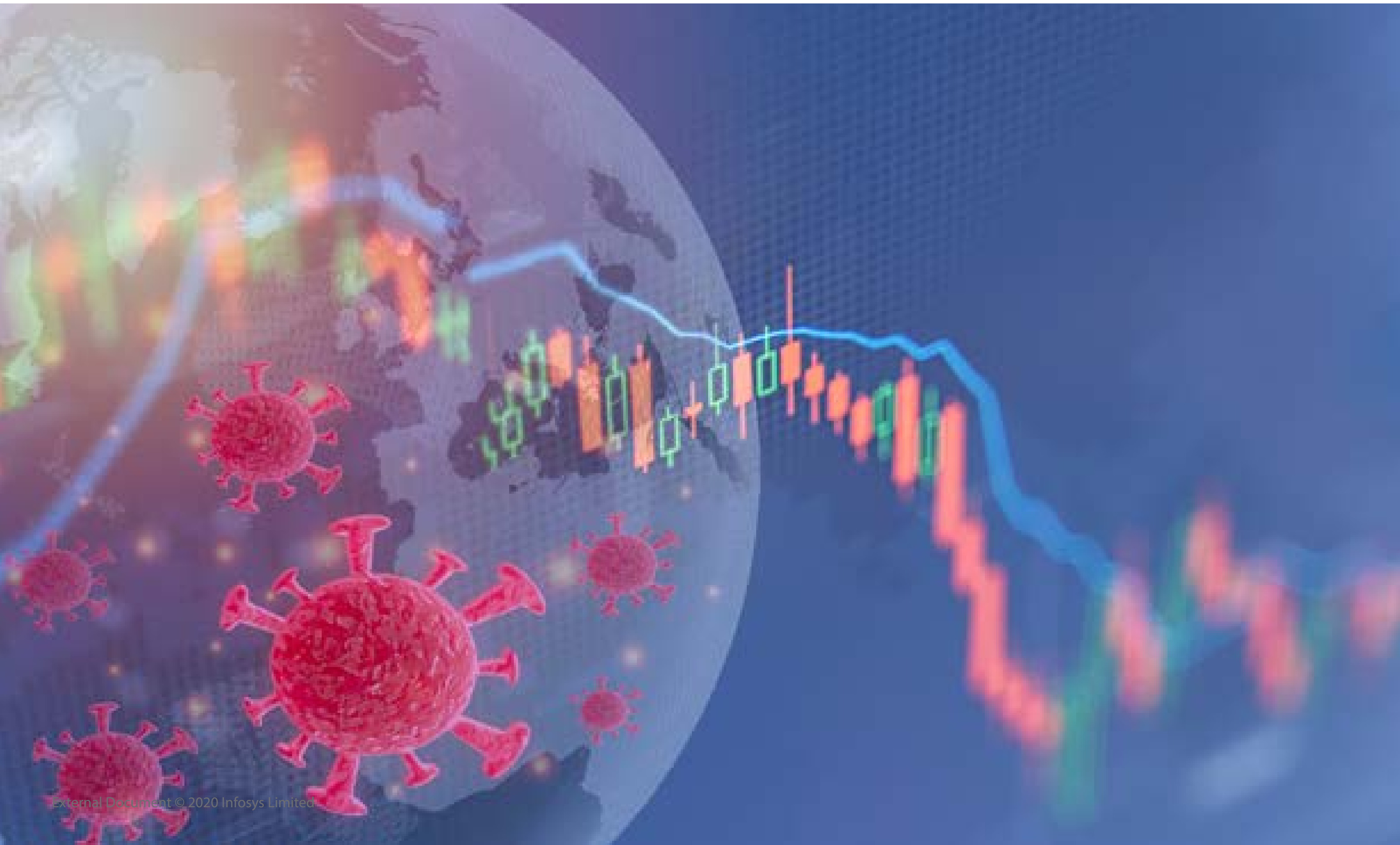
Traditional approaches in Epidemiology

- Data from surveys, investigators, hospitals
- Poor at prediction
- Costly
- Time consuming
- Dependable

Big Data Analytics in Epidemiology

- Usage of unstructrued data (like social media)
- Real time
- At a cheaper cost
- Needs standarsization
- Data could be inaccurate and biased

Traditional epidemiology uses data from surveys, investigators, hospitals, etc. Deriving information from such sources requires time and manpower.[2] It is reliable information but it is not real time, is slow and poor at prediction.[8] We live in a world where data is generated daily and there are systems which capture this data and analyze it. But these systems are parallel and there is no data integration and interoperability.[5] This need of data integration was realized by Taiwan and despite its proximity to China, Taiwan managed to contain the spread of COVID – 19 by creating big data for analytics by integrating its national health insurance database with its immigration and customs database, which helped them to create real time alerts.[16] The combination of advanced real time data analytics and integration of data from non-traditional data sources like social media, wearables, mobile health applications can identify and track outbreak as they are happening.[2] Conversations on social media platforms, search histories, trending hashtags, location tracking, aggregated data from health applications and wearables have the potential to detect an outbreak and contain it too.



Although Big Data Analytics would enable us to be better prepared for outbreaks in the future, it comes with its own set of challenges:

- While sharing data to enable big data analytics for picturing community's health and creating timely alerts for outbreaks is vital, it needs strong cybersecurity as data could be stolen or corrupted and someone could put false triggers in the systems and create an alert that an outbreak is happening when it is not.[2] For every positive signal for an outbreak, there may occur dozens of false positive ones which would require investigations and reducing this signal to noise ratio is going to be challenging.[10]
- Data collected from personal devices and social media has the potential of being misused. Profiling of individuals for commercial purposes can have negative consequences at an individual as well as societal level.[11]
- Using unstructured and informal data sources would be a faster means to detect an outbreak, but the data gathered from these sources has a huge probability of being inaccurate and biased.
- The legal and ethical issues of using social media data for formal disease surveillance is an area that needs attention.[10]
- Data collected from various unconventional sources is not standardized.[10]
- Even though the accessibility of internet is far better than what it was even a couple of years back, we need improvements in the current digital infrastructure to improve disease surveillance.
- Data practitioners would be overwhelmed with the amount of data that needs to be worked with for disease surveillance.
- We need enhanced interdisciplinary collaborations among people who understand analytics, computing, AI, ML, biology, pharmacology, biomedical implications and population models to leverage Big Data Analytics as a tool to fight pandemics.[7]

Conclusion

Timely detection of an outbreak can prevent it from becoming a pandemic. To prevent the lag in detection of an outbreak and monitor its progress, we need to aid traditional epidemiology with Big Data Analytics. We need platforms to be developed to support collaborations among companies as well as we need platforms for deidentified patient level data at global level to manage the outbreak as it progresses. We need analysis of unstructured data collected from unconventional sources to be integrated with structured data to enhance data interoperability and therefore detect patterns to create alerts for an outbreak, design protocols to treat the infection, analyze the success of techniques that could help in controlling the spread of infection, and thus respond to an evolving outbreak. We require robust digital infrastructure which helps gather data from every nook and corner of the world, standardizes it and at the same time does not create false positive alerts for an outbreak. This needs to be supported by laws, regulations and systems which enables consents for data to be fetched at an individual level and prevents its use for commercial purposes.

Currently, we are leveraging upon our ability to quickly build solutions around Big Data Analytics to fight against COVID -19. Some of these solutions are open to all, some are implemented at regional level and some at national level. But while we are doing this, it is important to realize that the ultimate goal should be to build an efficient ecosystem, which would be beneficial at a global level i.e. which would prevent any future outbreak to become a pandemic. Big Data Analytics, along with partnerships among stakeholders, has the potential to derive data from various sources and connect the dots to prevent and manage any future pandemics.

References

1. Bill Gates, We're not ready for the next epidemic We're not ready for it. But we can get there, March 18, 2015, https://www.youtube.com/watch?v=6Af6b_wyiwI
2. Big Data and Global Health, Big Data and Analytics for Infectious Disease Research, Operations, and Policy: Proceedings of a Workshop. National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Global Health; Forum on Microbial Threats. Washington (DC): National Academies Press (US); Dec 8, 2016
3. 6 Ways Pharmaceutical Companies are Using Data Analytics to Drive Innovation & Value
4. Cosgriff, C.V., Ebner D.K., Celi L.A., Data sharing in the era of COVID-19, Published May 2020, The Lancet, Digital Health, VOLUME 2, ISSUE 5, E224, MAY 01, 2020
5. 'Big data analytics will be the backbone for combating pandemics', Interview - Dr. Tanica Lyngdoh, May 8, 2020, https://www.ha-asia.com/big-data-analytics-will-be-the-backbone-for-combating-pandemics/
6. Google news
7. Kent J., Understanding the COVID-19 Pandemic as a Big Data Analytics Issue, April 02, 2020
8. From chaos to coherence: Managing pandemics with data CAN DATA ANALYTICS PREVENT FUTURE PANDEMICS?, by The Economist
9. History of 1918 Flu Pandemic
10. Opportunities and Challenges for Big Data and Analytics, Big Data and Analytics for Infectious Disease Research, Operations, and Policy: Proceedings of a Workshop. National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Global Health; Forum on Microbial Threats.Washington (DC): National Academies Press (US); Dec 8, 2016
11. Garattini C., Raffle J., Aisyah D.N., Sartain F. & Kozlakidis Z., Big Data Analytics, Infectious Diseases and Associated Ethical Impacts, Published August 24, 2017
12. COVID-19 Story Tip: Johns Hopkins Helps Lead Creation of National Covid-19 Database for 'Big Data' Studies, June 30, 2020
13. A public data lake for COVID-19 research and development
14. Tribble M.H., Cheung D., COVID-19 public datasets: our continued commitment to open, accessible data, August 13,2020
15. Sullivan L., Verizon Media Launches COVID-19 Search Engine For Research, April 8, 2020, Search and Marketing Performance Daily
16. Duff-Brown B., How Taiwan Used Big Data, Transparency and a Central Command to Protect Its People from Coronavirus, March 3, 2020

Author

Aparna Ekande

Senior Associate Consultant – Life Sciences, Infosys Ltd.

Infosys®
Navigate your next

For more information, contact askus@infosys.com