Introduction

Currently, there are more than 2 million CoVID-19 cases and 180 thousands deaths worldwide, and the pandemic has caused severe economic disruption and overburdened the healthcare system. Researchers around the world are working to understand the characteristics of the coronavirus and develop effective treatments. Since the pandemic progressed rapidly with many unpredictable changes, the focus of the researchers had changed from time to time to develop different strategies at different stages of the pandemic. In this project, we explored how research priorities shift as CoVID-19 progresses. We hypothesized that the initial efforts mainly concentrated on understanding the origin and characteristics of SARS-COV-2, and then focused on transmission patterns, clinical symptoms, developing treatments such as vaccines, with the latest effort concentrated on non-pharmaceutical interventions such as social-distancing. We believe that understanding this trend can shed light on the efforts that researchers have been undertaking to combat CoVID-19. This project can serve as a starting point for understanding how the development of COVD-19 affected research priorities and scientific development over time. Studies could compare the current research approach for coronavirus with previous efforts to combat Zika and Ebola outbreak. Furthermore, it could help researchers to be more prepared for the next potential zoonotic disease outbreak.

Dataset and Methods

The dataset comes from COVID-19 Open Research Dataset Challenge (CORD-19), an open Kaggle dataset prepared by the White House and a coalition of leading research groups. The original metadata contains 52365 unique scholarly articles about coronaviruses, and 18 columns on paper title, authors, abstract, published date, journal, etc. After filtering for articles with non-missing abstract, published in 2020, and containing keywords covid-19 and its synonyms, we are left with 2592 articles. We came up with 8 research categories and lists of associated keywords for each category. The categories are virus origin, transmission, risk factors, medical care, diagnostics and surveillance, vaccines and therapeutics, ethical and social science considerations, and non-pharmaceutical interventions. To extract the main focus of research articles, we string matched the abstract section with our list of keywords, and assigned the article 1 if a match is found and 0 otherwise.

Discussion

We noticed a general research focus shifting from finding cues to preventive measures. During January of 2020, papers on the origin of the virus dominated the literature with papers studying the transmission and diagnostic methods of the virus being of subsequent interest. This trend held throughout February; as more cases were reported worldwide, more efforts focused on intervention, diagnostics and surveillance, suggesting the importance of preventative measures. In Figure 2, changes in research trend corresponded with government and WHO policies; stay at home orders and lockdowns have led to an increase in surveillance and medical care research. The trend suggests that governments and researchers alike are no longer looking for a cure for the virus either due to failed trials or a lack of further options available. Instead this may also explain the shift to intervention type studies. If the infection cannot be cured then the next best option would be to develop preventative measures. However, this mindset also demonstrates that government officials and researchers alike are less invested in preventing those infected from dying and more interested in preventing further infections. That ideal may also explain the constant interest in transmission that has been prevalent throughout 2020.

Reference

COVID-19 Open Research Dataset Challenge (CORD-19). Retrieved April 22, 2020, from https://kaggle.com/allen-institute-for-ai/CORD-19-research-challenge/data

Taylor, D. B. (2020, April 21). A Timeline of the Coronavirus Pandemic. *The New York Times*. https://www.nytimes.com/article/coronavirus-timeline.html

Supplemental Information:

Our research categories were drawn from the Tasks section of this Kaggle data challenge. The lists of keywords we used for each category are listed below:

- Virus genetics, origin, and evolution

virus_origin = ['origin', 'genome', 'whole genomes', 'sequencing', 'spill-over', 'reservoir']

- Transmission, incubation, and environmental stability

transmission = ['transmission', 'incubation', 'asymptomatic shedding', 'environmental stability', 'contagious', 'prevention', 'disease models']

- COVID19 risk factors

risk_factors = ['risk factor', 'co-infection', 'pre-existing', 'preexisting', 'co-existing', 'coexisting', 'smoking', 'co-morbidities', 'high-risk patient groups', 'susceptibility']

- Medical care

medical_care = ['surge capacity', 'surge medical staff', 'personal protective equipment', 'ppe', 'nursing facilities', 'long term care facilities', 'nursing homes', 'shortages']

- Diagnostics and surveillance

diagnostics_surveillance = ['diagnostics', 'surveillance', 'screening', 'testing', 'mitigation measures', 'prevention', 'early detection', 'ELISA', 'rapid bed-side tests', 'PCR', ‘CRISPR']

- Vaccines and therapeutics

therapeutics = ['therapeutics', 'vaccines', 'drug development', 'animal model', 'viral inhibitors']

- Ethical and social science considerations

social_ethical = ['ethical', 'ethical considerations', 'ethics', 'novel ethical issues', 'secondary impacts', 'psychological', 'psychological health', 'stigma', 'fear', 'anxiety', 'misinformation','social media']

- Non-pharmaceutical interventions

inventions = ['non-pharmaceutical interventions', 'school closure', 'travel ban', 'social distancing', 'social-distancing', 'cost and benefit', 'mass gatherings', 'funding']

[Figure 1] During January of 2020, papers on the origin of the virus dominated the literature with papers studying the transmission and diagnostic methods of the virus being of subsequent interest. This trend held throughout February; however, as March approached interests were shifted and research on diagnostics techniques held priority while papers on the origin of the virus declined in priority. Although April has not concluded, it appears likely that research on diagnostics has once again dominated government, medical, and scientific interest while transmission and viral origins continue to decline in interest. Part of this shift may be attributed to the extent of research that can be conducted on the origins of the virus, for example, the viral genome can only be sequenced a limited number of times and repeated studies on the viral genome are not only redundant but likely provide no new information that warrant a publication. It also appears that papers surrounding the medical care infrastructure also did not warrant much interest at the beginning of 2020; however, based upon the current trend it appears as if research interest in the area is growing. This could be due to the deteriorating medical infrastructure that has been observed globally as government

officials and researchers alike have a vested interest in obtaining a better understanding of what has and has not worked with current medical infrastructure.

Whereas papers studying viral origins has dropped from accounting for nearly 45% of publications to just 20%, the percentage of papers on diagnostics has increased from 20% to account for nearly 25% and it currently the predominant focus for most research alongside studies on the virus's transmission which has retained a steady 25% share of research since January. While almost no studies were performed on intervention techniques at the start of 2020, the share of papers has grown to account for nearly 10% of all papers published. Based on trends it seems likely that the focus on interventions will only grow. This is likely due in part to the current lack of treatments available for the virus that often leaves medical staff with their hands tied and only able to offer symptomatic care without any real cure or treatment. Interestingly, the research focus appears to have shifted away from the search and development of therapeutics. Whereas the share of papers focused on therapeutics accounted for 10% of research papers in January by March it was only 5% of research papers and appears to be only getting smaller while other research areas take priority. This trend suggests that governments and researchers alike are no longer looking for a cure for the virus either due to failed trials or a lack of further options available. Instead this may also explain the shift to intervention type studies. If the infection cannot be cured then the next best option would be to develop preventative measures. Akin to the eradication of smallpox, rather than attempting to treat and develop a cure, a vaccination would be more beneficial. However, this mindset also demonstrates that government officials and researchers alike are less invested in preventing those infected from dying and more interested in preventing further infections. That ideal may also explain the constant interest in transmission that has been prevalent throughout 2020. Two rapidly growing areas of research are both the risk factors as well as social ethical implications of this deadly disease. While they accounted for almost no papers in January, by March and April they have accounted for nearly 10% of papers each. Once again, the focus on risk factors demonstrates the interest in preventing further spread of the coronavirus rather than looking for a cure. However, the increased interest in the social ethical developments surrounding the coronavirus opens perhaps another window to think about disease. In many ways this disease has forced the hand of many governments with the implementation of strict laws and orders which calls into question the role of government when faced with a large scale natural disaster such as a pandemic.