COVID-19: Understanding the incubation and contagious periods of the disease.

Ayomide Owoyemi[1], Vipul Dhariwal[2], and Lydia Tse[2]

***Abstract*—The epidemic control measures of COVID involve contact tracing, isolation, and quarantining. Having a good understanding of the epidemiology of the disease is important in guiding these public health measures to detect, control and prevent the spread of the disease. The aim of this study is to determine the incubation period and contagious period of COVID-19. We utilized the Semantic Scholar COVID-19 Open Research Dataset (CORD-19) which, at the time of completing this project, contains over 280,000 COVID-19 related literature. Additionally, we used a literature knowledge graph provided by Steenwinckel et al. that was derived from the CORD-19 data. The results of our study showed a mean incubation period of 8.41 days (SD = 5) with a weighted mean of 8.38, and a contagious period of 5.43 days (SD= 2.12) with a weighted mean of 5.38. While these results might not be reliable, they give a significant insight based on the number of studies involved in arriving at the estimates. There will be a need for a more robust way to analyze this large amount of data to arrive at a more reliable estimate of both periods.**

***Index Terms*—Knowledge graphs, Page Ranks, Dataset,Graphs, COVID-19**

I. Introduction

COVID-19 is presently the biggest public health challenge globally, and since it was declared a global pandemic it has resulted in over sixty million confirmed infections and more than a million deaths across all continents of the world. (COVID-19 map). While a vaccine for the disease is presently in development, the control of the epidemic has involved the use of non-pharmaceutical approaches like lock-downs, closure of public spaces, travel restrictions etc. [Nicola et al.2020], [Ferguson et al.2020]

These epidemic control measures also involve contact tracing, isolation, and quarantining. [Kucharski et al.2020] Having a good understanding of the epidemiology of the disease is important in guiding these public health measures to detect, control and prevent the spread of the disease. This involves knowing the incubation period and contagious period of the disease which will help in making effective quarantine and isolation decisions that will be helpful in controlling the spread of the disease. [Lauer et al.2020]

The aim of this study was to determine the incubation period and contagious period of COVID-19.

II. Methods

We utilized the Semantic Scholar COVID-19 Open Research Dataset (CORD-19) which, at the time of completing this project, contains over 280,000 COVID-19 related literature [COR2020]. Additionally, we used a literature knowledge graph provided by Steenwinckel et al. that was derived from the CORD-19 data [Steenwinckel et al.2020].

A. Extracting Relevant Literature

Out of all the papers and their texts, we had to find sentences that mentioned the incubation period. But, the search was not straight forward as searching for just the incubation period could have false positives like mentions of incubation period without any subsequent mentions of the number of days. We used different parameters and keywords using regex to find the sentences that are relevant to our search, an approach previously used by Savova et al. [Savova et al.2010]. We applied the same method to find the contagious period as well. We also used the Spike power tool [Goldberg2020] provided by CORD-19.

B. Preparing the Literature Knowledge Graph

While attempting to utilize the N-Triples of representing the knowledge graph using the rdflib Python library, we discovered that there were many invalid N-Triples due to invalid syntax. We resolved these errors by correcting and preserving as many N-Triples as possible. Thus, the resulting knowledge graph contained 5,151,961 N-Triples with 1,362,044 subjects and 9 predicates. The subjects represented varying aspects of the available CORD-19 literature including Digital Object Identifiers (DOIs), authors, citations, and other relevant paper metadata.

For our project objectives, we focused solely on the citation information of a paper since papers that are cited more often are likely to be more credible. Hence, as we will discuss later, the citation information provided by the knowledge graph was necessary for running the PageRank algorithm.

C. Generating the Citation Subgraph

To run the PageRank algorithm [Xing and Ghorbani2004], we first needed to generate a subgraph of relevant literature and citation information. Because we are primarily concerned with cited papers, we furthered refined the set of literature retrieved from the extraction process by using the literature knowledge graph to determine if the paper had been cited based on the DOIs. Papers that have yet to be cited were removed. Thus, our resulting subgraph only contained the citation predicate and the subjects were the DOIs of the relevant papers.

D. Running the PageRank Algorithm

We used the PageRank algorithm to determine the quality of the relevant literature. Thus, the quality of the literature using the PageRank algorithm was determined through counting the

[1]PhD Student at Department of Biomedical and Health Informatics,
[2]MS Student at Department of Computer Science

number of times a paper was cited and outputting the probability distribution to represent the likelihood that randomly selecting a citation from a paper would lead to any particular paper. We ran the algorithm on our citation subgraph using the NetworkX Python library.

E. Extracting Incubation/Contagious Period

Once we got the csv file with sentences we cleaned some basic symbols. Then we applied the CDQA library [Ozyurt, Bandrowski, and Grethe2020] and found the pos tags of sentences. Then using the NER model [Colic and Rinaldi2019] and CDQA library we extracted the number of days from each sentence (of each paper). And, after manual inspection we found the abnormalities and extremities and cleaned them again. Once we got number of days suggested by each paper, we multiplied it with its pagerank to get a weighted average.

III. Results

A. Incubation Period

From the initial knowledge graph data set of 280,000 research papers, we filtered 90,000 papers mentioning contagious period but after through cleaning for Incubation period of COVID-19 we got 8357 unique papers.

Average number of days of Incubation period suggested by all the paper came out to be 8.41 days with a standard deviation of 5 making the range from 3.5 to 13.5 days. The median came out to be 9 days. The weighted mean was 8.38.

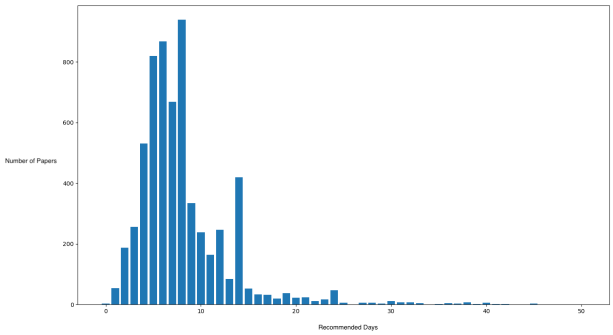

Fig. 1. Graph showing average number of days of Incubation Period recommended by number of papers

B. Contagious Period

Search for contagious period was more difficult as fewer papers mentioned it and we also had to be careful about including different keywords like infectious period. After cleaning the data set, we got 2700 papers mentioning Contagious Period of COVID-19.

Average number of days of contagious period suggested by all the paper came out to be 5.43 days with a standard deviation of 2.12 making the range from 3 to 7 days. The median came out to be 5 days. The weighted mean was 5.38.

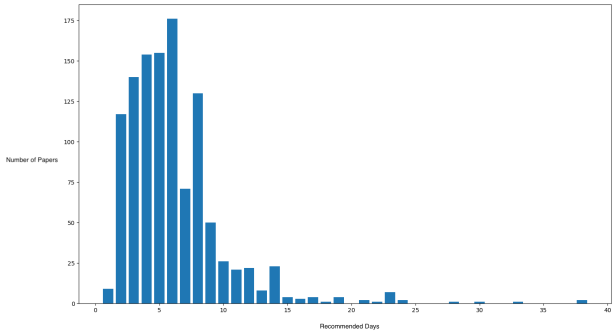

Fig. 2. Graph showing average number of days of Contagious Period recommended by number of papers

IV. Accuracy

We randomly took 100 papers and noted their suggested incubation period and another random 100 for contagious period, the average incubation period we got was 8.5 days which is very close to 8.8 that we found and the average contagious period came out to be 4.5 days which is also close to 5.2 that we found.

V. Discussion

Presently, the World Health Organization uses an incubation period of 2 – 14 days to inform its quarantine advice for member nations. The United State Center for Disease Control just revised its quarantine recommendations to 10 days without any symptom reported and without any need for testing required based on results of modelling they conducted. Having a good insight into the incubation and contagious period of COVID-19 is important in making public health decisions around quarantining and isolation for exposed and infected individuals which are necessary to controlling the spread of the disease.

The results of our study showed a mean incubation of 8.41 days (SD = 5) with a weighted mean of 8.38, this result differs from those obtained in systematic reviews conducted by Wassie et al., McAloon et al. and Quesada et al. which showed mean incubation periods of 5.7 days, 5.8 days, and 5.6 days respectively. [Wassie et al.2020], [McAloon et al.2020], [Quesada et al.2020] We also found a contagious period of 5.43 days (SD= 2.12) with a weighted mean of 5.38, this is different from the outcomes of studies conducted by Singanayagam et al. and Morone et al. which showed a period of 10 days. [Singanayagam et al.2020], [Morone et al.2020]

Out study has certain limitations that may affect the reliability of our results; the included studies were a mix of observational studies and randomized control trials, we used an aggregate of the incubation and contagious periods directly from the collected studies without conducting any meta analyses, we did not consider sexual or age differences in our estimates. We also did not consider the severity of the

disease in our analysis of the contagious period after illness from COVID-19.

However, we believe that our results give a valuable insight into the possible range of the incubation and contagious period for the COVID-19 disease based on the number of studies used for our estimates.

VI. Conclusion

The incubation contagious period of COVID-19 based on this study were 8.38 and 5.38 respectively. These results might not be reliable due to some the limitations highlighted, but they give a significant insight based on the number of studies involved in arriving at the estimates. There will be a need for a more robust analysis of this large amount of data to arrive at a more reliable estimate of both periods.

References

[Colic and Rinaldi2019] Colic, N., and Rinaldi, F. 2019. Improving spaCy dependency annotation and PoS tagging web service using independent NER services. *Genomics Inform.* 17(2).

[COR2020] 2020. CORD-19 | Semantic Scholar. [Online; accessed 4. Dec. 2020].

[Ferguson et al.2020] Ferguson, N.; Laydon, D.; Nedjati Gilani, G.; Imai, N.; Ainslie, K.; Baguelin, M.; Bhatia, S.; Boonyasiri, A.; Cucunuba Perez, Z.; Cuomo-Dannenburg, G.; Dighe, A.; Dorigatti, I.; Fu, H.; Gaythorpe, K.; Green, W.; Hamlet, A.; Hinsley, W.; Okell, L.; Van Elsland, S.; Thompson, H.; Verity, R.; Volz, E.; Wang, H.; Wang, Y.; Walker, P.; Walters, C.; Winskill, P.; Whittaker, C.; Donnelly, C.; Riley, S.; and Ghani, A. 2020. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. [Online; accessed 4. Dec. 2020].

[Goldberg2020] Goldberg, Y. 2020. SPIKE-CORD | AI2 Blog. *Medium*.

[Kucharski et al.2020] Kucharski, A. J.; Klepac, P.; Conlan, A. J. K.; Kissler, S. M.; Tang, M. L.; Fry, H.; Gog, J. R.; Edmunds, W. J.; CMMID COVID-working group; Emery, J. C.; Medley, G.; Munday, J. D.; Russell, T. W.; Leclerc, Q. J.; Diamond, C.; Procter, S. R.; Gimma, A.; Sun, F. Y.; Gibbs, H. P.; Rosello, A.; van Zandvoort, K.; Hué, S.; Meakin, S. R.; Deol, A. K.; Knight, G.; Jombart, T.; Foss, A. M.; Bosse, N. I.; Atkins, K. E.; Quilty, B. J.; Lowe, R.; Prem, K.; Flasche, S.; Pearson, C. A. B.; Houben, R. M. G. J.; Nightingale, E. S.; Endo, A.; Tully, D. C.; Liu, Y.; Villabona-Arenas, J.; O'Reilly, K.; Funk, S.; Eggo, R. M.; Jit, M.; Rees, E. M.; Hellewell, J.; Clifford, S.; Jarvis, C. I.; Abbott, S.; Auzenbergs, M.; Davies, N. G.; and Simons, D. 2020. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *Lancet Infect. Dis.* 20(10):1151–1160.

[Lauer et al.2020] Lauer, S. A.; Grantz, K. H.; Bi, Q.; Jones, F. K.; Zheng, Q.; Meredith, H. R.; Azman, A. S.; Reich, N. G.; and Lessler, J. 2020. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann. Intern. Med.*

[McAloon et al.2020] McAloon, C.; Collins, Á.; Hunt, K.; Barber, A.; Byrne, A. W.; Butler, F.; Casey, M.; Griffin, J.; Lane, E.; McEvoy, D.; Wall, P.; Green, M.; O'Grady, L.; and More, S. J. 2020. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. *BMJ Open* 10(8):e039652.

[Morone et al.2020] Morone, G.; Palomba, A.; Iosa, M.; Caporaso, T.; De Angelis, D.; Venturiero, V.; Savo, A.; Coiro, P.; Carbone, D.; Gimigliano, F.; Iolascon, G.; and Paolucci, S. 2020. Incidence and Persistence of Viral Shedding in COVID-19 Post-acute Patients With Negativized Pharyngeal Swab: A Systematic Review. *Front. Med.* 7.

[Nicola et al.2020] Nicola, M.; O'Neill, N.; Sohrabi, C.; Khan, M.; Agha, M.; and Agha, R. 2020. Evidence based management guideline for the COVID-19 pandemic - Review article. *Int. J. Surg.* 77:206–216.

[Ozyurt, Bandrowski, and Grethe2020] Ozyurt, I. B.; Bandrowski, A.; and Grethe, J. S. 2020. Bio-AnswerFinder: a system to find answers to questions from biomedical texts. *Database* 2020.

[Quesada et al.2020] Quesada, J. A.; López-Pineda, A.; Gil-Guillén, V. F.; Arriero-Marín, J. M.; Gutiérrez, F.; and Carratala-Munuera, C. 2020. [Incubation period of COVID-19: A systematic review and meta-analysis]. *Rev. Clin. Esp.*

[Savova et al.2010] Savova, G. K.; Masanz, J. J.; Ogren, P. V.; Zheng, J.; Sohn, S.; Kipper-Schuler, K. C.; and Chute, C. G. 2010. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J. Am. Med. Inform. Assoc.* 17(5):507–513.

[Singanayagam et al.2020] Singanayagam, A.; Patel, M.; Charlett, A.; Bernal, J. L.; Saliba, V.; Ellis, J.; Ladhani, S.; Zambon, M.; and Gopal, R. 2020. Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. *Eurosurveillance* 25(32):2001483.

[Steenwinckel et al.2020] Steenwinckel, B.; Vandewiele, G.; Rausch, I.; Heyvaert, P.; Colpaert, P.; Simoens, P.; Dimou, A.; Turkc, F. D.; and Ongenae, F. 2020. Facilitating covid-19 meta-analysis through a literature knowledge graph. In *Accepted in Proc. of 19th International Semantic Web Conference (ISWC)*.

[Wassie et al.2020] Wassie, G. T.; Azene, A. G.; Bantie, G. M.; Dessie, G.; and Aragaw, A. M. 2020. Incubation Period of Severe Acute Respiratory Syndrome Novel Coronavirus 2 that Causes Coronavirus Disease 2019: A Systematic Review and Meta-Analysis. *Curr. Ther. Res. Clin. Exp.* 93.

[Xing and Ghorbani2004] Xing, W., and Ghorbani, A. 2004. Weighted pagerank algorithm. In *Proceedings. Second Annual Conference on Communication Networks and Services Research, 2004.*, 305–314.