

코로나19 국제협력연구 분석 사례

유승찬 전문의
아주대학교
의료정보학과



1. 들어가는 글

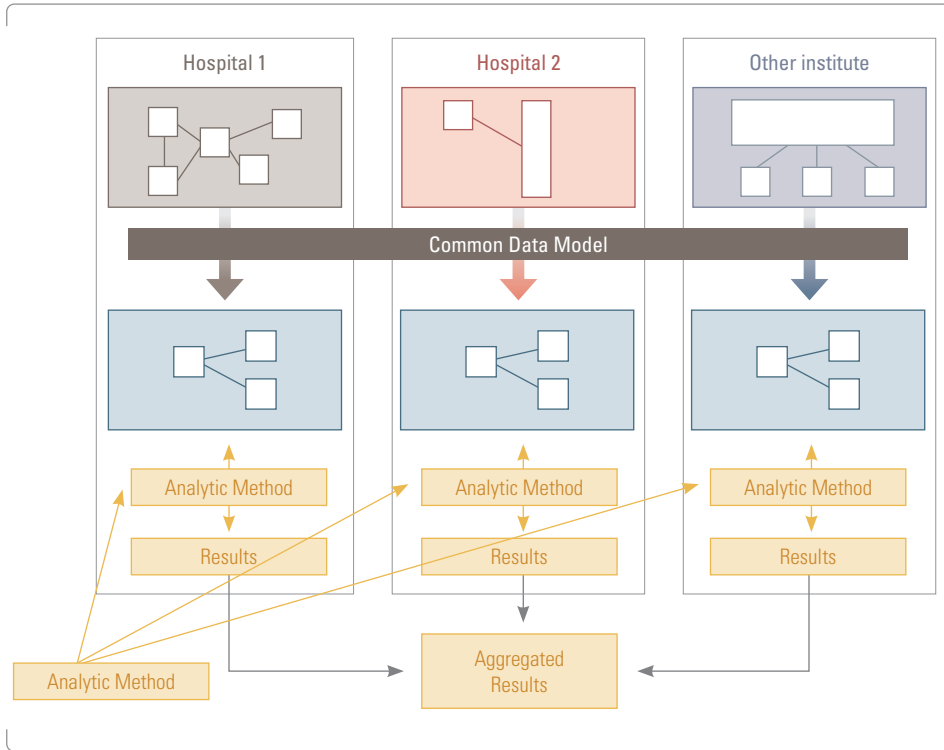
중국이 코로나바이러스감염증-19(Coronavirus disease 2019, 줄여서 코로나19)의 발병을 인정한 후 얼마 되지 않은 2020년 1월 20일 국내에서 코로나19 첫 확진자가 발생했다. 우리나라는 2020년 2월 21일 누적 환자 수 200명을 돌파하며 중국과 다이아몬드호 다음으로 코로나19 환자가 제일 많은 나라가 되었다. 이날 중국을 제외한 전 세계의 신규 확진자는 127명이었으며, 이 중 100명이 대한민국에서 발생했다. 당시까지 미국의 확진자는 15명, 영국 9명, 이탈리아 3명, 스페인 2명 등이었다[1]. HIRA 빅데이터 브리프 앞선 호에서 언급했다시피 2020년 3월 27일, 건강보험심사평가원(이하 심평원)에 축적된 전 국민 기반 보건의료빅데이터를 국내외 연구자들에게 제공하는 ‘코로나19 국제협력연구’가 시작되었다. 보건복지부(이하 복지부)와 심평원은 ‘코로나19 국제협력연구’라는 온라인 플랫폼을 구축하여 국내외 연구자에게 차별 없이 데이터를 활용할 수 있는 기회를 제공하여, 코로나19의 전 세계적 확산에 공동으로 대응할 수 있는 발판을 마련하였고, 전 세계적 위기상황에 우리나라의 리더십을 보여주었다.

2. 분산연구망, 공통 데이터 모델과 오딧세이 컨소시엄

분산연구망(distributed research network)이란 공통데이터모델(common data model, CDM)을 기반으로, 각 기관의 데이터를 표준화한 후에 데이터를 한곳에 모아서 분석하지 않고, 분석 프로그램을 각 기관에 보내어 기관 내에서 분석된 결과값만을 취합하여 연구를 수행하는 방식을 말한다. 외부 연구자가 기관 내의 데이터를 직접 보지 않고 연구를 수행할 수 있기 때문에,

민감정보 사용에 의한 개인정보 침해 우려가 적고 다국적 다기관 연구를 더욱 수월하게 진행할 수 있다는 장점이 있다.

[그림 1] 분산연구망의 개념도



오딧세이(Observational Health Data and Informatics, 이하 OHDSI) 컨소시엄은 OMOP-CDM 기반의 분산 연구망을 이끄는 국제적인 비영리 연합체이다. 전 세계 산학연 기관들이 자유롭게 참여하고 활동하고 있으며, 비교효과분석, 의약품질검증, 약물 안정성 평가, 경제성 평가 등 보건의료 빅데이터를 활용하여 새로운 의학 근거를 창출하는 연구를 수행하고 있다. 강력한 오픈소스 정책으로, OMOP-CDM 기반의 보건의료데이터의 분석을 위한 다수의 분석툴을 개발하고 배포하고 있기도 하다.

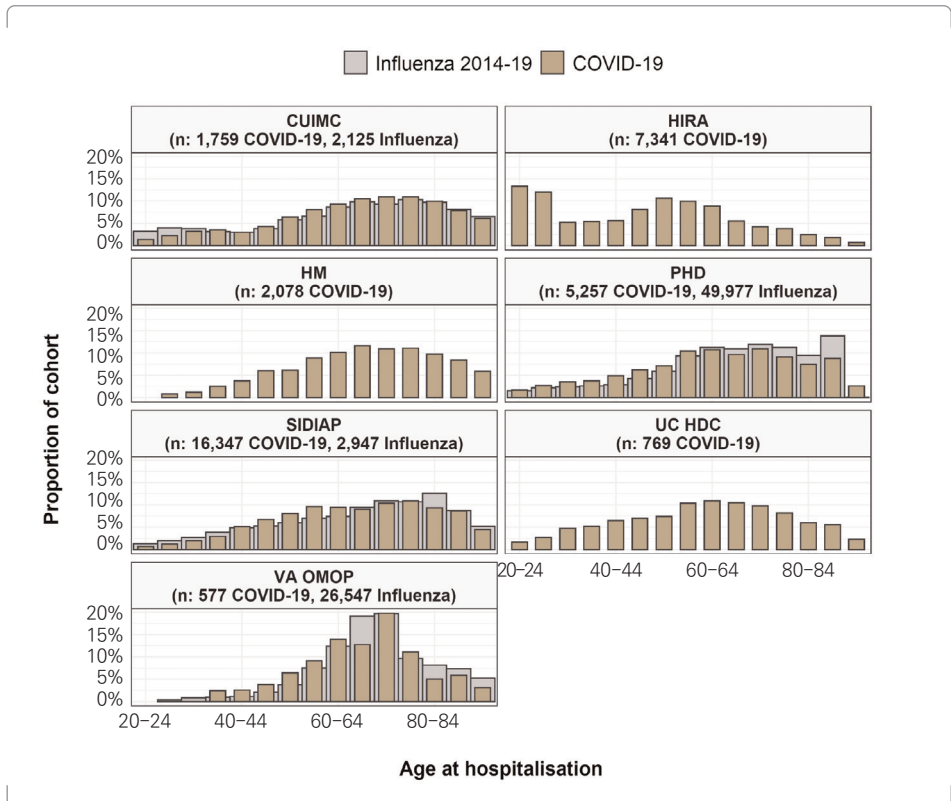
오딧세이 컨소시엄은 2020년 3월 26일부터 28일까지, 3일간 행사를 시작으로 코로나19 연구마라톤(study-a-thon, 이하 스터디톤)을 심평원의 '코로나19 국제협력연구'와 병렬적으로 진행하였으며, 본고에서 그 연구 결과를 소개하고자 한다.

3. 전 세계 코로나19 입원 환자 특성 분석

심평원 코로나19 데이터를 활용한 국제협력 연구로 전 세계 3만 4천여 명의 코로나19 환자의 특성을 분석한 첫 번째 심층 표현형(Deep phenotyping) 연구이다[2]. 코로나19로 입원한 한국 환자 7천여 명, 미국 환자 8천여 명, 그리고 스페인 환자만 8천여 명을 포함한 총 3만 4천여 명의 환자의 분산 연구망 데이터 시스템을 활용하여 국가별/기관별 코로나19 환자들의 특성을 분석하였다. 코로나19가 발생하고, 단일 기관이나 단일 국가에서의 코로나19 환자들의 특성을 분석한 연구가 주요 저널에 많이 발표되었지만, 본 연구를 통해서 기관별, 나라별 코로나19 환자들의 특성이 매우 다를 수 있음을 확인하였고, 단일 기관 및 단일 국가 데이터만을 이용한 코로나19 연구 결과를 일반화할 때 주의를 기울여야 함을 이해할 수 있었다.

다른 나라와 비교하여 한국 코로나19 입원 환자의 특성은 비교적 나이가 어리고, 여성 비율이 높다는 특성이 있었다. 코로나19 사태 비교적 초반의 연구 결과로, 신천지 사태로 감염자 중에 젊은 여성이 비교적 많았고, 높은 의료 접근성으로 한국은 당시 대부분의 코로나19 환자를 입원시킬 수 있었기 때문에 이러한 차이가 보였으리라 여겨진다.

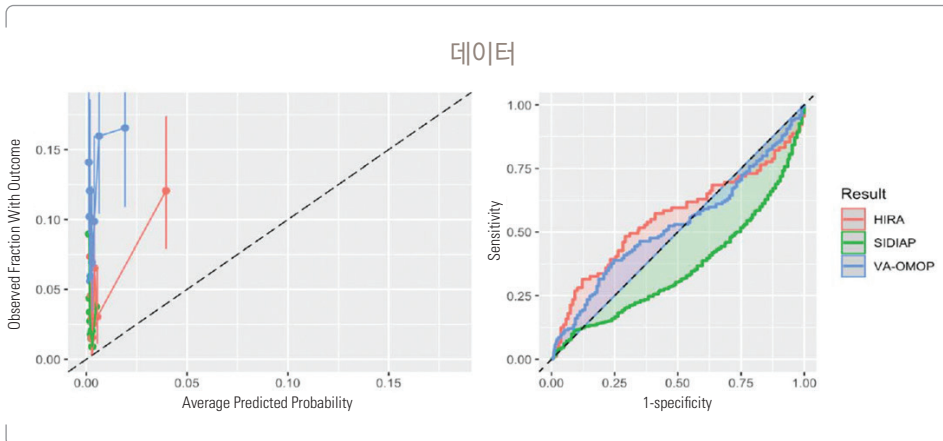
[그림 2] 코로나19로 입원한 환자들과 독감으로 입원한 환자들의 나이 분포



4. 코로나 19 중증 위험환자 예측 모형

코로나19의 발생과 더불어 다양한 연구가 발표되었으며, 입원해야 할 정도로 중증 경과를 보일 환자를 예측하는 모형에 관한 연구들도 많이 있었다. 그중에서 C-19 지표(COVID-19 vulnerability [C-19] index)가 많은 주목을 받았다. 오딧세이 컨소시엄에서는 해당 지표를 한국 심평원, 미국, 스페인의 코로나19 환자 데이터베이스를 이용해 검증해 보았고, 모형의 변별력을 측정하는 ROC 곡선 면적(AUROC, the area under a receiver operating characteristics curve)을 측정하였을 때, 0.36에서 0.56 정도로 매우 낮음을 확인하였다[3]. 보통 AUROC가 0.7 이상일 때 예측 모형이 일정 수준의 예측력을 갖고 있다고 평가한다.

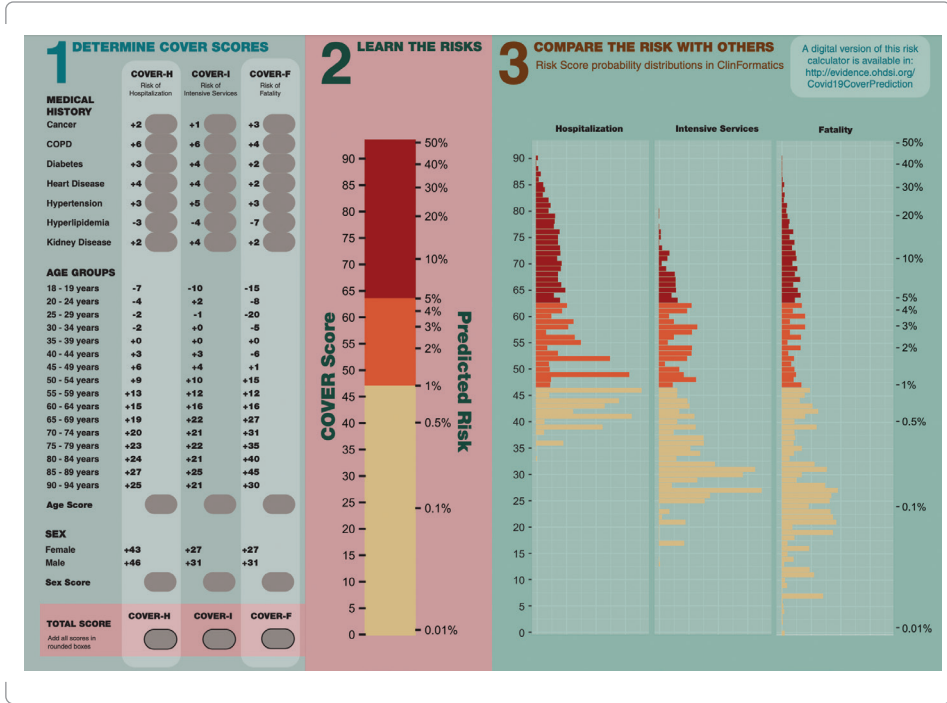
[그림 3] 코로나19 환자 대상 C-19 예측력 확인을 위한 ROC 곡선 및 calibration plot.



(데이터원) HIRA: 한국 청구 데이터; SIDIAP: 스페인 청구데이터; VA-OMOP 미국 Veterans Affairs 병원 의무기록

이에 오딧세이 컨소시엄에서는 다국적 다기관 데이터 네트워크를 활용하여 COVER (COVID-19 Estimated Risk) 점수를 개발하였다. COVER 점수는 독감 환자의 데이터를 활용해 개발되었고, 나이, 성별 및 기저질환을 통해 감염병에 취약한 환자들을 예측하고자 하였다. 이 예측 모형은 한국 심평원 자료를 비롯하여 스페인, 미국의 5개 데이터베이스, 약 4만 4천 명의 코로나19 환자 데이터를 통해 검증되었다. 코로나19 감염 시 입원할 확률, 중환자실 입실 확률, 그리고 사망확률을 예측하는 COVER-H, COVER-I, 그리고 COVER-F의 예측력(AUROC)는 각각 0.69-0.81, 0.73-0.91, 0.72-0.90으로 우수한 것으로 나타났다[4].

[그림 4] COVER 점수 계산기



주: 나이, 성별과 기저 질환을 통해 점수를 계산할 수 있으며 코로나19 감염 시 COVER-H는 입원할 확률, COVER-I는 중환자실 입실 확률, COVER-F는 사망 확률을 각각 예측함

5. 고찰

현재 세계는 코로나19라는 전대미문의 재난에 허덕이고 있다. 이 재난은 국가의 보건 의료 시스템뿐 아니라 교육, 경제, 외교, 산업 시스템 전반에 막대한 영향을 끼치고 있다. 재난 영화의 장면들이 우리의 주변에서 재현되고 있으며, 우리는 코로나19가 세계에서 사라진다고 하더라도 우리가 받은 흉터와 유산은 사라지지 않으리라는 것을 알고 있다. 하지만 지금 이 상황에서 무엇보다 걱정되는 것은 코로나19가 과연 사라질 것이냐는 의문이다. 유례없이 빠른 속도로 백신이 개발되고 있고, 많은 전문가들은 2021년 안으로는 백신의 승인과 상용화가 가능할 것으로 예상하고 있다. 하지만, 질병의 박멸(disease eradication)은 조절(control), 또는 질병의 제거(disease elimination)과는 다른 개념이다. 질병의 박멸(eradication)은 전 세계로 질병의 발생이 한 케이스도 생기지 않는 개념인 데 반해, 질병의 제거(elimination)은 지속적인 지역 사회 감염 및 타지역으로의 감염이 없는 상태를 뜻한다. 현재까지 인류가 백신을 이용하여 박멸(eradication)한 감염병은 천연두 (smallpox)가 유일하다. 1980년대 천연두의 박멸을 선언한 후에, WHO는 소아마비(polio)의 전 세계적 박멸(global eradication)과 홍역의 지역적 제거(regional elimination)을 위하여 큰 노력을 기울이고 있다. 두 질병 모두 오랜 기간 동안 개발된 백신이 존재하지만, 아직 두 질병

모두 해당 목표에는 도달하지 못하였고, 도달할 수 있을지에 대해서도 의문이 많다. 코로나19 바이러스의 숙주가 인간뿐이 아니라는 점과 무증상 감염자의 존재를 생각한다면, 전 세계 박멸(global eradication)이 아니라 지역적 제거(regional elimination) 또는 조절만도 아주 쉽지 않은 목표일 것이다. WHO의 홍역에 대한 지역적 제거의 정의는 낮은 발생률, 높은 수준의 감시(surveillance)와 감염 발생에 대한 빠른 대처(outbreak) 그리고 높은 인구 집단 면역 상태이다. 현재 세계는 코로나19 박멸(eradication)보다는 제거(elimination)와 같은 목표를 먼저 생각해봐야 할 것이다[5].

코로나19에 대한 백신만으로 질병이 박멸(eradication)되지 않고, 소규모 감염이 국지적으로 지속한다면, 우리는 어떤 준비를 해야 할까? 8월 말 한국의 코로나19 상황은, 질병의 제거와 같은 이상적인 상황보다는 당장 폭발적으로 급증하는 환자들을 어떻게 치료해야 할지에 대한 현실적인 고민부터 해야 할 것이다. 주저하고자 하는 사실은, 백신 개발까지만 버티는 대책은 매우 위험할 수 있다는 점이다.

또 하나의 희망은 코로나19 치료제의 개발이다. 하지만 현재까지의 결과는 다소 실망스럽다. 많은 사람들이 기대했던 말라리아 치료제인 하이드록시클로로퀸(hydroxychloroquine)과 항바이러스제인 칼레트라 lopinavir/ritonavir)는 임상시험 결과 치료제로서의 뚜렷한 혜택을 보여주지 못하였다. 특히 하이드록시클로로퀸은 심장 부정맥을 증가시킬 수 있어, 선부른 약물의 사용이 얼마나 위험할 수 있는지 보여주었다[6]. 스테로이드는 중증 환자에서 사망률을 낮추어 주었지만, 사망률을 24.6%에서 21.6%로 낮추는 정도였고, 렘데시비르는 회복 기간을 15일에서 11일로 단 4일 줄여주는 데 그쳤다. 많은 기대를 모았던 혈장치료 역시 임상시험에서 뚜렷한 효과가 없었다[7]. 폐렴 및 저산소증, 그리고 동반되는 다양한 합병증에 대한 보존적 치료(supportive care) 및 후유증 예방이 현재까지는 코로나19 치료에 가장 주요한 치료 방침 중 하나이다.

코로나19에 대한 백신 또는 치료제가 나오기만 한다면, 이 사태가 모두 끝날 것으로 생각하는 많은 사람이 있다. 하지만 감기, 홍역, 독감 등의 다양한 바이러스 질환을 보면 백신 개발이 어려울 수도 있고, 백신이 개발되어도 질병의 완전한 박멸이 어려울 수도 있고, 효과적인 백신과 치료제가 있더라도 지속적인 감염과 이로 인한 사망이 생길 수도 있다는 점을 간과해서는 안 된다. 막연한 희망보다는 상황 악화의 가능성을 염두에 두고 대책을 마련해야 할 것이다.

결국 코로나19 감염 시 중증으로 이환될 환자들을 먼저 예측하고, 이러한 취약 계층에 대한 보다 면밀한 방역 대책이 필요할 것이다. 또한, 현 코로나19 환자들의 치료 패턴과 동반 합병증에 대한 보다 많은 데이터가 필요하다. 한국에서 코로나19 대규모 전파가 발생한 지 6개월이 지난 이 시점에서, 코로나19 환자들의 중장기 후유증에 대한 데이터가 있어야, 이를 예방 또는 치료할 수도 있을 것이다. 우리는 아직 코로나19라는 질병에 대해 아는 것보다 모르는 것이 훨씬 많기 때문에 지속적인 연구가 필요하다.

이러한 점에서 심평원의 코로나19 국제 협력연구 플랫폼 개발 및 착수는 매우 중요한 의의를 지닌다. 코로나19 발생 당시 많은 의료 전문가와 정부 관계자들은 청구데이터보다는 의료진들이 환자의 상태와 치료를 기록한 레지스트리(registry) 데이터에 주목했다. 실제로 청구자료에는 진단이 부정확한 경우도 있고, 증상에 대한 자료가 없기 때문에 연구에 필요한 많은 자료가 부족하다. 하지만, 코로나19 환자 치료 때문에 격무에 시달리는 의료진들에게 레지스트리 데이터 작성을 요구하는 것은 무리가 따르며, 실제로 시의적절하게 데이터가 구성되지도 못했다. 재난이 닥치면, 대응하는 모든 인력은 평소보다 훨씬 많은 업무량을 처리해야 한다. 이럴 때 새로운 데이터를 만드는 것도 물론 필요하지만, 기존에 활용 가능한 데이터를 먼저 확인하고, 이를 활용할 방안부터 검토하는 것이 얼마나 중요한지 확인할 수 있었다.

또 한 가지 괄목할만한 부분은 온라인 플랫폼 및 분산연구망을 활용하여 한국 및 전 세계 연구자들이 데이터를 활용할 수 있도록 했다는 점이다. 위에서도 얘기했듯이 분산연구망을 활용한 연구는 원칙적으로 연구자들이 해당 데이터를 보지 않고도 연구를 진행할 수 있기 때문에, 개인 정보 침해의 위험이 매우 낮다. 당시 한국의 코로나19 환자들의 신상정보가 많이 노출된 상황에서 환자들의 10년간의 의료 이용 기록을 연구자에게 직접 개방하는 것은 많은 문제를 야기할 수 있었다. 분산연구망을 활용하지 않았다면, 이렇게 빠른 시기에 데이터를 활용할 수는 없었을 것이다. 또한 연구자들의 차별을 두지 않아, 많은 전 세계 젊은 연구자들이 활발하게 참여할 수 있었다. 데이터를 일부 국내의 저명한 학자들에게만 제공했다면, 전 세계 젊은 연구자들의 참여가 많이 제한되었을 것이다. 이번 프로젝트를 통하여 대한민국 의료 시스템 및 심평원의 위상이 제고되었고, 코로나19 연구에서의 국제적 리더십을 보여주었다.

6. 맺음말

심평원 코로나19 데이터를 활용한 코로나19 국제협력연구 사례 분석을 통해, 이번 프로젝트가 한국의 국제적 위상 제고 및 코로나19 연구에 어떤 기여를 했는지 고찰했다. 다국적 다기관 공동 협력 연구를 통해서, 우리는 단일 국가 및 기관의 코로나19 환자 데이터가 얼마나 편향될 수 있는지, 그리고 이를 이용한 예측 모형의 예측력이 높지 않음을 확인할 수 있었다. 코로나19 사태를 통해 우리는 다시 한 번 팬데믹에 대한 국제적 공조의 중요성을 확인했고, 본 프로젝트를 통해 일정 부분 한국도 이러한 전 세계 노력에 기여했음에 매우 자랑스럽게 생각한다. 시의적절한 데이터를 통한 신종 감염병의 이해와 감시의 중요성을 이해하고, 부처를 나누지 않고 정부와 민간의 데이터 융합을 가속하고, 많은 연구자가 활용할 수 있게 하는 방향을 위해서도 계속 노력해야 할 것이다.

참고문헌

- [1] WHO. Coronavirus Disease (COVID-19) Situation Reports [Internet]. [cited 2020 Aug 24]. Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>
- [2] Burn E, You SC, Sena A, Kostka K, Abedtash H, Abrahao MTF, et al. Deep phenotyping of 34,128 patients hospitalised with COVID-19 and a comparison with 81,596 influenza patients in America, Europe and Asia: an international network study. medRxiv. 2020 Jun 28;2020.04.22.20074336.
- [3] Reps JM, Kim C, Williams RD, Markus AF, Yang C, Salles TD, et al. Can we trust the prediction model? Demonstrating the importance of external validation by investigating the COVID-19 Vulnerability (C-19) Index across an international network of observational healthcare datasets. medRxiv. 2020 Jun 17;2020.06.15.20130328.
- [4] Williams RD, Markus AF, Yang C, Salles TD, Falconer T, Jonnagaddala J, et al. Seek COVER: Development and validation of a personalized risk calculator for COVID-19 outcomes in an international network. medRxiv. 2020 May 29;2020.05.26.20112649.
- [5] Heywood AE, Macintyre CR. Elimination of COVID-19: what would it look like and is it possible? Lancet Infect Dis [Internet]. 2020 Aug 6 [cited 2020 Aug 24];0(0). Available from: [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(20\)30633-2/abstract](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30633-2/abstract)
- [6] Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB. Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review. JAMA. 2020 May 12;323(18):1824-36.
- [7] Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. JAMA [Internet]. 2020 Jul 10 [cited 2020 Aug 24]; Available from: <https://jamanetwork.com/journals/jama/fullarticle/2768391>