

감염병의 세계적 유행에 대응하는 빅데이터 활용

Response to COVID-19 using healthcare big data in Korea



히윤정 국회의원

Key Points

- ☑ 2020년 3월 11일, 세계보건기구는 코로나19의 세계적 유행(pandemic) 선포, 각국 정부는 코로나 19에 효과적 대응 위한 정보 제공 중
- ☑ 현재 세계보건기구 및 각국 정부가 제공하는 정보는 코로나19 확진자와 사망자 관련 발생 현황 및 이동 정보를 시각화하는 방식
- ☑ 오픈사이언스 개념 도입으로 역학, 의료이용 등 다양한 정보로 구성된 빅데이터 활용, 국제사회와 협력으로 인류 편익 도모

Key Words

코로나19, 감염병, 빅데이터, 오픈사이언스, 공동데이터모델(CDM), 국제 협력
COVID-19, pandemic, big data, open science, common data model(CDM), international cooperation

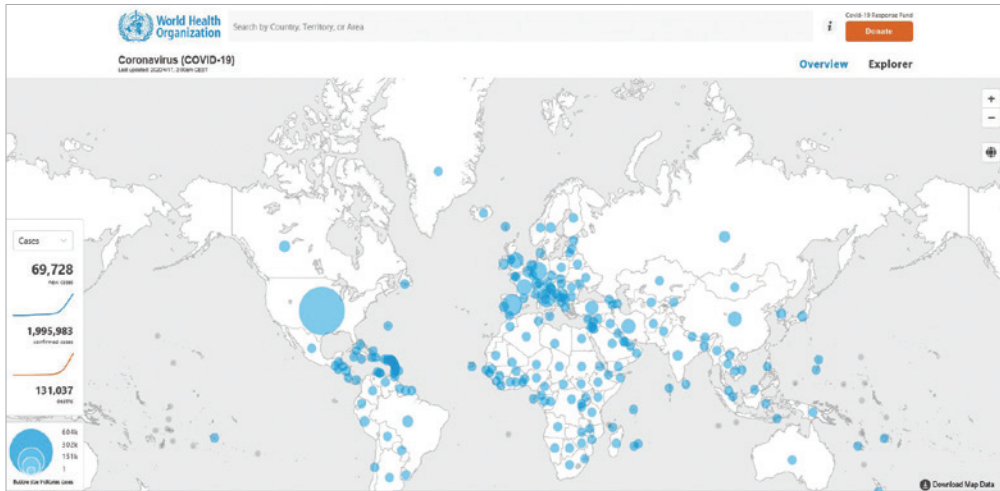
1. 감염병의 세계적 유행(pandemic)

2020년 3월 11일 세계보건기구(World Health Organization, 이하 WHO) 사무총장은 신종 코로나바이러스 감염증(Novel Coronavirus Disease 2019, COVID-19, 이하 코로나19)을 세계적 유행(pandemic) 단계¹⁾로 선포하였다. 2019년 12월 초 중국 후베이성의 우한시에서 처음 발생 및 보고된 후 2020년 4월 17일(오전 9시 기준) 현재 전 세계 216개국에서 2,091,786명의 환자가 발생하였고, 이 중 142,000명이 사망하여 6.78%의 치사율²⁾을 보이고 있다. 한국에

1) WHO의 전염병 경보는 6단계로 나누어져있으며, 1단계는 동물 사이에 한정된 전염, 2단계는 야생 동물과 가축 사이에 전염되다가 소수의 사람에게도 전염 가능성이 있는 단계, 3단계는 동물 사이 및 동물과 사람 사이에 전염이 발생하였으나 공동체 수준의 감염이 이루어지지 않은 단계, 4단계는 사람 사이에 전염이 이루어지며 공동체로 확대되어 세계적 유행의 가능성이 있는 단계(epidemic), 5단계는 전염이 확대되어 한 권역(대륙) 내 2개국 이상에서 전염이 확대된 단계, 6단계는 국가-대륙 간 전염이 이루어져 세계적 유행이 진행되는 단계로 5, 6단계를 팬데믹(pandemic) 단계라고 칭한다.

2) 치사율 = (총 사망자수 ÷ 총 확진자수) × 100

서도 2020년 1월 20일 중국에서 입국한 중국 여성이 첫 확진 판정을 받았고, 2월 20일부터 대구·경북 지역을 중심으로 환자가 큰 폭으로 증가하여 8,236명의 확진자가 발생하였으며, 이 중 81명이 사망함으로써 치사율은 0.98%이다(그림 1).



[그림 1] 전 세계 코로나19 발생 현황

자료: World Health Organization. Novel Coronavirus (COVID-19) Situation(4.17. 제네바 시간대 기준)

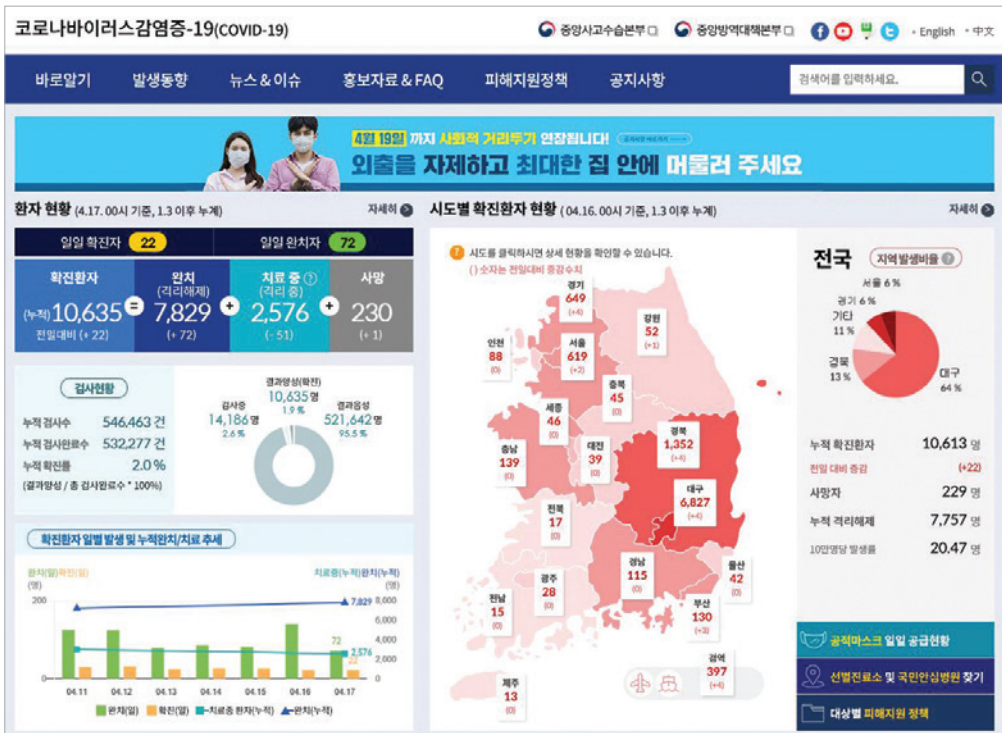
중국 우한시에서 처음 확진자가 발생한 이래 약 100일 만에 아시아를 넘어 유럽과 미주 대륙으로 확진자 수가 증가하는 추세를 보임으로써, 각국 정부는 자국민의 지역 간 이동 통제, 국경 폐쇄, 특정 항공노선 운항 금지, 입국 금지 등 다양한 정책을 동원하여 가능한 한 감염경로를 차단하려는 노력을 기울여 왔다. 그러나 이러한 봉쇄(containment)의 방침에도 확산 속도와 규모가 잦아들지 않자, WHO는 결국 감염병의 세계적 유행을 선언하게 되었고, 이후 방역정책은 점차 봉쇄에서 완화(mitigation) 기조로 전환되어 사회적 거리두기(social distancing), 휴교령, 집단 행사 취소 등 감염의 속도를 늦추는 정책이 한동안 지속될 것으로 보인다.

2. 데이터 활용 시각화 정보 제공

감염병의 세계적 유행은 유행병이 특정 지역에 국한되지 않고 범대륙적 또는 전 세계적으로 유행하는 상태를 말한다(WHO, 2020). 14세기 중세 유럽 인구의 3분의 1인 최대 2억 명 이상의 사망이 발생했던 흑사병이나 천연두 같은 유행도 있지만, 2009년 발생했던 신종 인플루엔자(Swine Flu, H1N1, 이하 신종플루)와 같은 독감류의 유행도 많은 사람들의 기억에

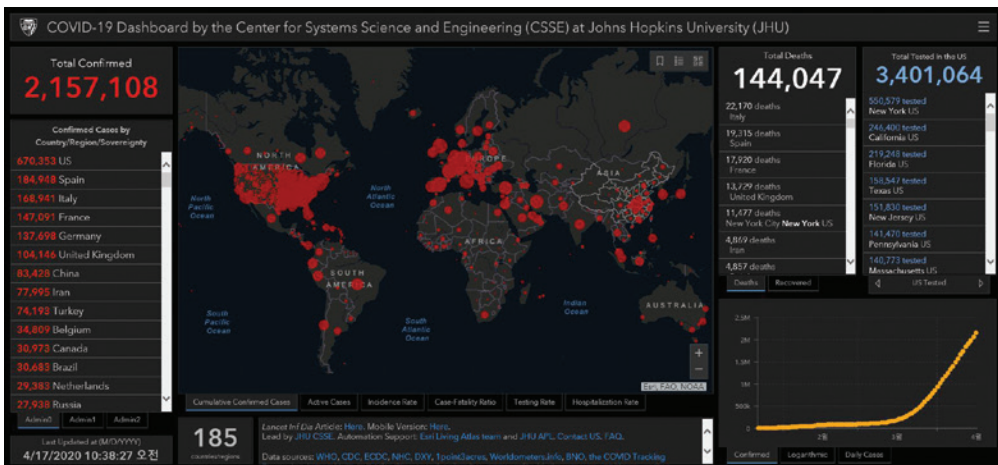
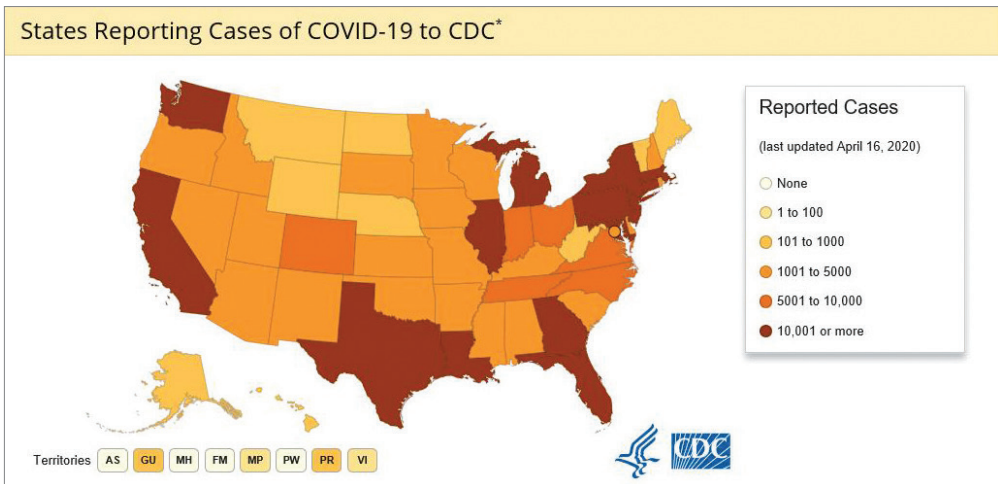
선명하게 남아 있다. 특히 독감은 1918년 1차 세계대전 중 발생한 스페인 독감을 비롯하여 아시아 독감, 홍콩 독감 등 비교적 단기간 내에 여러 유행병을 발생시킨 역사가 있다(World Economic Forum, 2020). 이번에 발생한 코로나19는 코로나 바이러스에 의해 촉발된 첫 번째 세계적 유행이지만 전염 패턴이 독감과 유사한 것으로 알려져 있으며, 특히 1960년대 후반 홍콩 독감과 2009년 신종플루에 이어 세 번째로 WHO가 감염병의 세계적 유행을 선언하고 있다는 점에서 주목을 끌고 있다. 전 세계적인 인적·물적 교류 활성화는 지구촌 구석구석을 촘촘히 연결해 주고 있으며, 이는 감염병 유행에 있어서도 예외일 수 없다.

따라서 세계의 여러 국가들은 매년 감염병의 유행을 예의 주시하고 검역 등 이에 대비한 프레임워크도 갖추어 왔다(Wang 등, 2020). 그리고 최근 발생하는 감염병에 대해 WHO 등 책임 기구 및 각국 정부가 대시보드 기능을 이용하여 실시간 확진자 현황 등을 시각화하여 일반 국민들이 알고 싶어 하는 정보를 적시에 제공하는 노력을 기울이고 있다(그림 2, 3). 특히 한국의 경우 정부의 투명한 정보 공개 방침을 토대로 지방자치단체별 확진자 동선을 공개 하면서 이를 민간이 활용하여 시각화 화면을 개발하는 활동도 이루어져, 정보통신 등 기술의 발전으로 과거에 비해 국민이 스스로 감염의 위험을 줄여갈 수 있도록 도움을 주고 있다.



[그림 2] 질병관리본부 코로나19 상황판

자료: 질병관리본부 홈페이지. <http://ncov.mohw.go.kr> (검색일자: 2020.4.17.).



[그림 3] 미국의 코로나19 환자 발생 상황판

자료: 미국 질병통제센터(CDC) 홈페이지. <https://www.cdc.gov/coronavirus/2019-nCoV/index.html>
 미국 존스홉킨스 Center for Systems Science and Engineering(CSSE) 홈페이지. <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6> (검색일자: 2020.4.17.)

그러나 감염병의 세계적 유행 시에는 환자 발생 및 이동 현황뿐 아니라 진료현장에서 임상의학의 진단과 환자 치료에 도움을 줄 수 있는 실제 임상자료 기반 근거와 정부의 정책 수행에 필요한 정보를 제공하는 것이 매우 중요하다고 할 수 있다. 이를 위해서는 정확하고 견고(robust)한 역학 데이터, 실제 진료를 통해 수집된 임상데이터, 기저질환을 파악할 수 있는 과거 의료이용 데이터 및 여행기록 데이터 등 다학제적 빅데이터 구축이 필요하다 (Wang 등, 2020; Xu와 Kraemer, 2020).

3. 실제임상데이터 기반 오픈 사이언스 구현의 필요성

최근 몇 년 사이의 컴퓨팅 하드웨어 성능과 인터넷 속도 등 디지털 기술의 발달과 확산은 국경과 지역을 넘어 과학기술 연구의 시스템적 성과 향상에 크게 기여하였다(신은정 등, 2017). 이러한 기술적 토대는 연구를 수행하는 과정에서 아이디어 및 데이터의 공유, 실시간 소통과 참여를 가능하게 하여 협력적 연구 수행이 가능한 환경이 조성되고 있다(신은정 등, 2017). 디지털 기술 기반의 연구 환경 변화를 바탕으로 OECD(2015)는 “공공연구의 성과물을 디지털 형태로 공개 및 확산시켜 사회경제적 편익을 제고하려는 일련의 노력”을 오픈 사이언스(open science)라 정의하고, 유럽 내에서는 이를 오픈 액세스(open access, 성과물 관련), 오픈 데이터(open data) 및 개방 협력(open collaboration)의 영역으로 확대하여 사회 공동체가 그 성과를 누릴 수 있도록 정책 수립을 권고하고 있다(EC, 2016; 신은정 등, 2017 재인용).

이러한 과학기술 연구 환경의 변화를 이번 코로나19 상황에 대입하여 보면, 현재 환자 발생 및 이동 경로 정보가 실시간으로 제공되는 수준을 넘어, 감염병의 특성을 이해하고, 환자의 상태를 호전시키며, 향후 예방 및 감시 방안을 모색할 수 있는 다양한 정보가 통합적으로 제공되고 활용될 수 있는 방안에 대한 고민이 필요하다.

현재 미국의 경우 코로나19 상황을 오픈 사이언스로 접근하도록 백악관이 직접 국립보건연구원(National Institute of Health, NIH)을 통해 코로나19 관련 기사나 논문 내용을 인공지능(AI) 방식으로 읽어 많은 연구자들이 그 내용을 공유할 수 있도록 데이터세트(COVID-19 Open Research Dataset, 이하 CORD-19)를 구축하여 제공하였다. 연구자들은 이 데이터세트를 이용하여 각자 모든 자료를 건별로 읽고 검토하지 않아도 내용을 파악할 수 있으며, 연구주제가 바뀌어도 다시 자료를 정리하지 않아도 될 수 있게 효율적 정보 공유 방식을 제시하였다. 또한 현재 코로나19에 대한 학술 논문은 비교적 단기간 내에 다수의 논문이 출판되고 있는데, NEJM, Lancet, JAMA³⁾ 등 의료 분야 주요 학술지에서도 근거 공유를 적극적으로 장려하여 동료심사(peer review) 후 게재가 아니라 사전 게재(preprint)라는 형식을 통해 별도의 서버에 먼저 게시하고, 이후에 동료심사 과정을 거쳐 정식 게재하는 파격적인 심사시스템을 운영하고 있다. 이러한 논문들은 모두 오픈 소스로 공개하여 구독료 없이 이용할 수 있도록 접근성을 높이고 있다.

이처럼 국제사회 내에 데이터를 공개하고 다수의 연구자 간 협업 필요성에 대한 공감대는 형성되어 있지만, 코로나19를 극복하고 문제를 해결하기 위한 임상근거는 아직 부족한 실정이다. 현재까지의 임상적 특징에 대한 이해는 대부분 개별 기관 중심(single center study)의 연구에서 나오기 때문에 선택편향(selection bias)의 이슈가 존재하고, 표본수가 통

3) NEJM : New England Journal of Medicine, JAMA : Journal of American Medical Association

계적 유의성을 확보하기에 매우 부족하거나, 중국 우한의 초기 단계 연구에서 나온 결과가 대부분이라는 제한점이 있어 이 결과들을 일반화하여 적용하기는 어려운 현실이다. 이러한 제한점을 극복하기 위해 환자의 구체적 임상 데이터를 활용하기 위해서는 민감한 개인정보의 활용이라는 더 큰 어려움을 극복해야 할 것이다.

4. 나가며

현재 지구상 대부분의 국가가 당면하고 있는 코로나19라는 세계적 유행의 감염병을 극복하기 위해서는, 앞서 언급한 다양한 측면의 검토가 필요하다. 우선 감염병을 이해하고 환자를 효과적으로 치료하며, 질병의 확산과 치명률을 낮추는 유의미한 연구결과 도출과 이의 시의적절한 활용을 위한 '데이터 확보'가 핵심적이라고 할 수 있다. 한국은 건강보험 제도를 통해 디지털 방식으로 전 국민의 의료이용 정보를 수집·관리하고 있다. 이는 다른 나라에 비해 양질의 데이터를 신속하게 활용할 수 있는 환경이라고 볼 수 있다. 따라서 환자의 기저질환 정보가 포함된 청구데이터를 역학 데이터(유전체 정보 포함), 병원의 EMR 데이터, 환자의 이동경로 데이터 등과 통합하여 분석할 수 있다면, 감염병의 감염성 및 위험도가 높은 인구집단에 대한 이해, 사망률과 잠복기를 포함한 감염의 자연력 확립, 원인 유기체의 식별 및 특성화, 효과적인 예방과 통제조치를 위한 역학 모델링 등 다양한 측면에서 임상 근거의 수준을 높일 수 있는 기반이 될 수 있다(McCall, 2020).

이를 위해서는 창의적이고 다소 파격적인 노력과 시도가 필요하다. 먼저, 전 세계적 감염병 확산 위험이 있는 경우, 공익을 위해 민감정보 간 결합·활용을 예외적으로 인정하는 내용이 포함된 개인정보보호법 시행령 개정 논의가 필요하다. 이 경우 민감한 개인정보가 유출되어 악용되는 것을 막을 수 있도록 공통데이터모델(Common Data Model, CDM)⁴⁾ 등의 방법을 적용하여 익명화한 후 사용하는 방법도 검토해 볼 수 있다. 이와 함께 현재 당면하고 있는 다양한 연구 질문⁵⁾들을 서로 다른 시각에서 고민하고 해결책을 찾아 볼 수 있도록 국제사회가 참여할 수 있는 협력연구 플랫폼을 구축하는 것도 필요하다. X

4) 공통데이터모델은 각 기관 별 상이한 데이터 용어와 구조를 표준화하여, 원데이터는 기관 내 보관하고 분석 코드만 공유한 후 결과값(근거)만 공유하도록 지원하는 익명화된 데이터 모델이다.

5) 코로나19 감염 이후 사망 및 합병증 발생 고위험군 선별, 기저질환과 사망의 상관관계, 신약 임상시험 대상군 선정 등

참고문헌

- 신은정, 안형준, 양현채, 최병삼, 양승우, 정원교, 김수연. 오픈사이언스정책의 도입 및 추진 방안. 과학기술정책연구원. 2017.
- European Commission. Open Innovation, Open Science, Open to the World. 2016.
- McCall B. COVID-19 and Artificial Intelligence: protecting health-care workers and curbing the spread. *Lancet Digital Health*. 2020. [https://doi.org/10.1016/S2589-7500\(20\)30054-6](https://doi.org/10.1016/S2589-7500(20)30054-6).
- OECD. "Making Open Science a Reality", OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing. 2015.
- Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. *JAMA*. 2020. DOI:10.1001/jama.2020.3151.
- World Economic Forum. A Visual History of Pandemics. <https://www.weforum.org/agenda/2020/03/a-visual-history-of-pandemics>. (검색일자: 2020.3.16.)
- World Health Organization. WHO Pandemic Phase Descriptions and Main Actions by Phase. https://www.who.int/influenza/resources/documents/pandemic_phase_descriptions_and_actions.pdf?ua=1 (검색일자: 2020.3.11.)
- Xu B, Kraemer MUG. Open Access Epidemiological Data from the COVID-19 Outbreak. *Lancet Infect Dis*. 2020. [https://doi.org/10.1016/S1473-3099\(20\)30119-5](https://doi.org/10.1016/S1473-3099(20)30119-5).