

코로나19 대응을 위한 국제협력연구의 현황과 전망

박정환 사무관 보건복지부 의료정보정책과
양광모 교수 삼성서울병원
정재훈 교수 가천대학교 의과대학
이상열 교수 경희대학교 의과대학
박혜정 교수 연세대학교 강남세브란스병원
노연숙 부장 건강보험심사평가원 빅데이터연구부

1. 들어가는 글

「코로나19 대응 국제협력연구」(이하 '코로나19 국제협력연구')에 참여한 연구자들의 경험 공유 및 의견수렴을 위한 간담회를 개최하였으며, 본 원고는 간담회에서 논의된 내용을 바탕으로 작성되었다. 간담회 참여자는 코로나19 국제협력연구 관계자 및 연구 참여자로 구성되었다.

2. 주요 내용

1) 「코로나19 대응 국제협력연구」의 의의



박정환 사무관

“코로나19 국제협력연구는 국가 간 경계를 두지 않고 국내외 연구자에게 보건의료 빅데이터를 활용한 연구 기회를 제공하여 전세계적 보건 위기에 주도적으로 대응하고자 시도되었습니다.”

코로나19 국제협력연구는 건강보험심사평가원(이하 심평원)에 축적된 전 국민 기반 보건의료빅데이터를 국내외 연구자들에게 제공하여 실제임상데이터(Real-world Data, RWD)를 활용한 시의적절한 임상 및 정책 근거 창출을 위해 시작되었다.

세계보건기구(WHO) 설립 이래 3번째 ‘감염병의 세계적 유행(Pandemic)’ 선언이 이루어질 정도로 유래 없는 전 세계적 확산이 이루어졌으나, 코로나19에 효과적으로 대응하기 위한 임상·정책 근거는 턱없이 부족하여 관련 학회와 방역당국을 중심으로 연구를 위한 임상데이터 요구도가 높았다. 특히 다른 나라에 비해 모범적 방역을 실시하고 있는 대한민국 사례를 자국에 적용하고자 하는 외국의 관심도 높아 다양한 경로로 협력 연구 요구도 증가하고 있었다.

이에 보건복지부(이하 복지부)와 심평원이 ‘코로나19 국제협력연구’라는 온라인 플랫폼을 구축하여 국내외 연구자에게 차별 없이 데이터를 활용할 수 있는 기회를 제공함으로써, 코로나19의 전 세계적 확산에 공동으로 대응하여 의미 있는 성과를 도출할 수 있는 계기가 될 것으로 기대하였다. 또한 심평원은 코로나19 전체 환자의 진료정보를 신속하게 수집·정제하여 개방할 수 있다는 장점이 있어 시의적절한 연구 지원이 가능하다는 것을 확인하는 계기도 되었다.

‘코로나19 국제협력연구’ 프로젝트는 글로벌 보건 위기에 전 세계적으로 연대하는 계기가 되고자 국내외 경계를 두지 않고 기획하였다. 이러한 형태의 협력연구는 처음 시도된 것으로 해결해 나가야 할 제한점들이 있지만, 주기적 데이터셋 업데이트를 통한 데이터 안정화 및 질병관리본부에서 제공하는 확진자 데이터와 청구데이터를 융합하는 등 지속적인 노력을 통해 보완점을 찾아가고 있다. 이러한 일련의 과정이 새로운 데이터 활용 모형을 구축하는 기반이 될 것으로 기대하고 있다.

2) 데이터 관련 이슈



“데이터셋 업데이트 과정에서 분석 결과의 변이를 줄이고 안정적이고 신뢰도 높은 데이터셋을 구축하는 방안에 대한 고민이 필요하다고 생각합니다.”

이상열 교수

코로나19의 지속적 확산으로 국제협력연구용 데이터셋은 일정 시점에 업데이트가 이루어지고 있으나, 기존 데이터셋을 이용하여 분석한 연구자가 이후 업데이트된 데이터셋으로 재분석한 경우 결과의 차이가 발생할 우려가 있었다. 실제로 업데이트된 데이터셋으로 분석을 시행 해보았을 때, 초기 제공된 데이터셋을 이용한 분석결과에서 보였던 몇 가지 의문점들이 업데이트된 데이터셋의 분석결과에서 상당히 해결되면서 데이터의 정확도와 안정성이 확보되어 간다고 평가되고 있다. 이는 코로나19 확진자 발생과 치료가 진행되고 있는 과정에서 제공하는 데이터셋의 특성 상 어쩔 수 없는 부분이 있지만 가능한 사전 공지와 설명을 통해 보완할 필요가 있다.



양광모 교수

“데이터 기초분석 결과를 함께 제시해 준다면 연구 가이드 역할이 될 것 같아 결과적으로 연구의 완성도를 높일 수 있지 않을까 기대합니다.”

다른 이슈로는, 같은 주제의 연구에 대해 연구자별로 상이한 결과가 나오는 부분이다. 물론 동일한 데이터를 활용하더라도 연구자의 연구 목적과 분석 관점 및 분석 기법에 따라 결과에 차이가 있을 수 있지만, 이번 연구 프로젝트의 경우 동일한 코호트를 활용하고 있으므로 좀 더 신중하게 접근할 필요가 있다. 이 문제를 해결하고 연구의 신뢰도를 높이기 위해 데이터 개방 주제(심평원)가 데이터의 성격과 경향을 보여줄 수 있는 baseline 분석 혹은 기저 질환 분석 결과를 함께 제공한다면, 연구자들이 연구 과정에서 분석의 방향성을 확인하고 완성도를 높이는데 도움이 될 수 있다는 의견이 있었다.

코로나19 발생 초기에는 의료진들이 데이터 입력보다는 환자를 살리는데 더 초점을 두다보니, 청구 데이터 뿐 아니라 의료기관의 임상데이터에서도 정제되지 않은 값들이 발견되는 경우가 있었다. 이는 감염병 발생(outbreak) 시점에 데이터를 개방하는 상황에서는 공통적으로 발생하는 문제점이라고 여겨지지만, 이런 오류 패턴을 모델링하는 방법도 연구해 볼 가치가 있을 것으로 생각한다. 향후 이런 상황이 발생할 때마다 데이터를 개방하겠다는 정책적 판단을 하고 있다면, 현재 마주한 고민이 매우 중요한 지식 기반이 될 수 있을 것이다.

3) 새로운 분석 환경에 대한 경험 공유



정재훈 교수

“오픈 소스 통계패키지 기반 분석이 가능한 환경이 마련된다면 보다 다양한 방법론을 활용한 연구가 가능할 것이라고 생각합니다.”

청구데이터를 이용한 분석 경험이 많아도 이번 코로나19 국제협력 연구 프로젝트의 운영 방식인 분석코드만 업로드하는 경험은 처음이라 ‘과연 데이터를 열람하지 않고 작성한 코드가 잘 돌아갈까’ 하는 불안감이 있었던 것도 사실이다. 평소에는 데이터의 기술통계량을 확인하면서 수정하는 방식으로 분석을 해왔는데, 이번 연구는 코드를 제출하고 결과를 받아서 수정하는 방식으로 진행하다 보니 시간이 좀 더 소요되는 것은 어쩔 수 없는 사실이다. 코로나19와 관련한 연구는 해외에서도 활발히 진행되고 있어서 연구가 얼마나 빠르게 진행되어서 결과가 나오느냐가 중요한데, 코드를 제출하고 피드백을 받는 과정을 거치다보니 직접 접근하여 분석하는 것 보다는 시간이 지체된 점은 아쉬운 부분이었다.

새로운 분석 방법과 함께, 연구자들의 다양한 니즈 및 오픈 소스 소프트웨어 기반 분석에 대응할 수 있는 유연하고 확장 가능한 분석 환경 조성의 요구도 진지하게 고민해야 할 시점이라고 생각한다. 심평원도 다양한 빅데이터 분석 기법과 이를 위한 분석패키지 적용을 위한 환경 개선에 관심을 갖고 여러 측면에서 방법을 고민하고 있으나, 현재 국가정보원이 허용한 공공기관의 공공데이터 개방 가이드라인 하에서는 제한이 많은 것이 사실이다. 또 다른 방법으로 서버 증설을 통해 더 많은 원격 계정 구축으로 폐쇄망에 직접 접근하여 분석하는 방법은 고려하고 있지만 단기간에 이루어지기는 어려울 수 있고, 특히 감염병 등 민감한 정보를 다루는 상황에는 여전히 제한점이 있을 수 있는 방법이다. 또한 표준화된 데이터로 정형분석이 가능하고 기관 간 데이터 이동 없이 개인정보를 보호하는 환경에서 다기관 분석이 가능한 공통데이터모델(Common Data Model, CDM) 기반 분석환경도 2021년까지는 심평원 내 환경에서 안정화 될 것으로 기대하고 있다. 코로나19 국제협력연구를 기회로 향후 다양한 기관이 보유한 데이터를 연계 융합하여 새로운 근거와 가치를 생산할 수 있는 분석 기반 마련을 위한 적극적 의견 수렴 등 다각도에서 고민하고자 한다.

4) 데이터 제공 방식 및 범위에 관한 답변



박혜정 교수

“현재 제공되는 건강보험심사평가원의 청구 데이터와 임상 정보와 같은 추가적인 데이터의 연계가 이루어진다면 지금보다 다양하고 확장된 연구가 가능하지 않을까요?”

코로나19 국제협력연구는 건강보험 청구데이터를 전문적으로 분석해 본 경험이 없는 연구자들이 접근하기에 다소 어려움이 있어, 익숙하지 않은 연구자들도 쉽게 접근할 수 있는 방안 마련이 필요하다는 의견이 있었다. 이는 본 연구방식으로 개방되는 건에만 해당되는 사안은 아니나, 기초적인 분석 가이드가 마련되어 제공된다면 큰 도움이 될 것으로 생각된다.

또한 심평원이 보유한 추가 자료(DUR, 적정성 평가 결과 등)와 요양기관의 투약시점, 증상발현 시기 등 임상데이터 연계를 통해 심평원 데이터의 장점을 살리고 활용가치를 극대화 할 수 있는 방안이 마련되면 좋겠는 의견이 있었다. 이는 하나의 데이터 제공 주체가 해결할 수 없는 현안이나, 최근 통과되어 시행을 앞두고 있는 데이터 3법 개정안의 가이드라인이 적용되면 어느 정도 보완될 수 있을 것으로 기대하고 있다.

본 연구프로젝트는 동일 코호트를 사용하는 환경에서 연구자의 연구아이디어 보호 차원에서 세부 연구주제 및 연구방법을 공개하지 않는 방식으로 진행되고 있으나, 연구주제 중복 등의 문제가 발생할 수 있기 때문에 추후에는 연구의 키워드만 공유하는 등 연구 성과와 효율성을 동시에 고려할 수 있는 방안이 마련된다면 더 발전적인 연구가 진행 될 수 있을 것이라는 의견이 있었다.

3. 결론

코로나19 국제협력연구는 현재 이슈가 되고 있는 문제의 해결에 도움을 줄 수 있도록 관련 데이터를 연구자들에게 신속하게 제공하고, 이를 국내외에 동시에 개방하여 국제협력연구 체계를 구축하고 새로운 빅데이터 활용 환경을 조성하였다는 데 의의가 있다. 이에 본 간담회는 감염병의 전 세계적 유행이라는 긴박한 시점에, 이전에 시도되지 않은 새로운 방식으로 프로젝트를 추진하면서 연구 참여자들의 경험을 공유하고 향후 보완해 가야 할 사안을 점검하는 의미 있는 기회가 되었다.

이번 간담회에서는 민감한 데이터의 즉시적 활용을 위해 물리적으로 매우 제한적인 환경에서 데이터를 이용해야 한다는 점과 더 풍성한 연구를 위한 임상 데이터 부재의 아쉬움이 주로 논의되었고, 데이터 업데이트 등 데이터의 안정화와 이를 통한 연구 전반의 신뢰도 향상 등에 관한 의견이 개진되었다. 현재 본 연구프로젝트를 포함하여 코로나19 관련 연구가 국내에서 활발하게 진행되고 있으나, 최근 해외에서도 대규모 확진자 데이터가 급속히 수집되면서 임상연구, 백신·치료제 개발연구 및 정책연구가 점차 정교하게 이루어지고 있다. 본 간담회 내용을 토대로 앞으로 국내 연구도 데이터 경쟁력을 강화하기 위한 다각적인 노력과 변화 시도가 지속적으로 이루어지는 계기가 될 것으로 기대해 본다.