Textbook for Korea Field Epidemiology Training Program

역학조사관 교육 표준교재

: 역학 역량 강화 과정

역학조사관 교육 표준교재 Ⅱ
: 역학 역량 강화 과정

첫째판 1쇄 인쇄 | 2025년 04월 25일
첫째판 1쇄 발행 | 2025년 05월 12일

지 은 이 한국역학회
출 판 기 획 김도성
편집디자인 김영준
표지디자인 군자출판사 표지부
일 러 스 트 군자출판사 일러스트부
발 행 처 질병관리청

© 2025년, 역학조사관 교육 표준교재 Ⅱ권: 역학 역량 강화 과정
이 책은 저작권법에 의해 보호를 받는 저작물이므로 무단 복제와 전재는 금지되어 있습니다.
단, 교육적 목적으로 복사하여 사용할 수 있으며, 상업적 목적이나
제품판매의 목적으로는 사용할 수 없습니다(비매품).

이 책은 질병관리청 정책연구용역사업 『역학조사관 교육 표준교재 개발』(주관연구기관: 한국역학회)을 통해서 개발되었습니다.

ISBN 979-11-7068-257-8
세트 ISBN 979-11-7068-255-4
발간등록번호 11-1790387-100448-01

발간사

안녕하십니까.

전 세계적인 유행으로 모두의 일상과 안전을 위협하고 사회 전반에 큰 피해를 안겨 주었던 코로나19를 경험하면서 역학조사의 중요성을 다시 한번 실감하였습니다. 신종감염병의 위기 속에서 국민의 건강과 안전을 지키기 위해 최전선의 현장에서 발로 뛰며 헌신한 이들이 바로 역학조사관입니다. 이들은 감염병의 특성을 알아내고, 확진자와 그 접촉자를 찾아 감염병의 전파를 막아냄으로써 우리 사회가 안정을 찾는 데 핵심적인 역할을 해왔습니다. 감염병이 발생하면 항상 현장으로 제일 먼저 달려가 묵묵히 힘든 업무를 수행해 온 역학조사관들에게 깊은 감사의 마음을 전합니다.

2025년은 우리나라에 역학조사관 제도가 도입된 지 25년이 되는 뜻깊은 해입니다. 2000년도에 질병관리청(당시 질병관리본부)은 공중보건의사를 중심으로 제1기 역학조사관 20명을 선발하고 감염병과 역학 등에 관한 교육을 시작하였습니다. 그리고 2015년 중동호흡기증후군(MERS)의 국내 유행 이후, 국가방역체계 개편과 함께 업무의 연속성과 전문성을 겸비한 공무원 중심의 제도로 전환하고 현재와 같은 역학조사관 교육·훈련의 틀을 마련하였습니다. 올해는 전문 역학조사관 제도를 도입한 지 10년이 되는 해이기도 합니다.

장기간 지역사회 전파가 지속되었던 코로나19 유행을 겪으면서 현장 대응의 중요성이 더욱 커졌고, 이에 따라 중앙과 시·도 외에 시·군·구까지 역학조사관을 의무적으로 배치하도록 했습니다. 현재 전국에 교육과정을 수료한 250여 명의 역학조사관이 활동하고 있으며, 역학조사관이 되기 위한 교육·훈련 과정에 있는 수습역학조사관도 약 300명에 달합니다.

그만큼 이들의 전문성을 키우고 역량 있는 인재를 양성하기 위한 교육이 더욱 중요해졌습니다. 특히 실무에 필요한 전문 지식과 현장에서 즉시 활용 가능한 경험을 체계적이고 일관성 있게 전달하기 위한 표준화된 교육 자료의 필요성이 매우 절실해졌습니다.

이번 『역학조사관 교육 표준교재』의 발간은 그 자체로서 의미가 큽니다. 표준화된 교육 교재 제공을 통해 역학조사관 제도가 한 단계 도약하는 계기가 마련되길 바랍니다.

총 3권으로 구성된 이 교재는 감염병의 이해와 관리 및 대응체계 등에 관한 기본적인 이론부터 역학조사에 필요한 전문적인 이론과 분석 방법, 그리고 다양한 현장 역학조사 사례까지 폭넓고 깊이 있게 담고 있습니다. 이 교재를 통해 역학조사관이 다양한 상황에서 효과적으로 대응하기 위해 알아야 할 핵심 이론과 실무 역량을 충분히 익힐 수 있을 것으로 기대합니다. 아울러, 역학조사관으로서의 첫걸음을 내딛는 이들에게 든든한 길잡이가 되기를 진심으로 희망합니다.

감염병은 언제든지 우리를 위협할 수 있습니다. 이에 대한 대비와 대응은 결국 '사람', 즉 현장을 책

임지는 전문가들에게 달려 있습니다. 이 교재가 역학조사관들이 현장에서 보다 빠르고 정확하게 업무를 수행하는 데 기본 자료로 적극 활용되기를 바라며, 나아가 국가방역체계를 더욱 공고히 하는 밑거름이 되기를 바랍니다.

끝으로, 교재 발간을 위해 애써 주신 한국역학회 천병철 회장님을 비롯한 집필진 여러분, 그리고 편집위원 등 모든 전문가분들께 깊이 감사드리며, 함께 참여해 주신 질병관리청의 직원분들께도 감사의 말씀을 드립니다.

2025년 5월

질병관리청장

지영미

서문

이 책은 역학조사관 기본교육 교재이다. 그러나 이 책은 기존의 기본교육 내용에 대한 강의록이 아니다. 그렇게 책을 만들면 아주 쉬웠을 것이다. 기존에 강의하는 강사들에게 강의 내용을 정리해서 달라고만 해도 교재가 되었을 것이다. 그렇지만 1년이 안되는 짧은 기간 동안 국내와 국외 역학조사관 학습목표와 역량에 대한 것을 모두 정리하고 40회가 넘는 회의를 진행하면서 역학조사관 학습목표를 모두 새롭게 개발하고 이에 맞추어서 교재를 완전히 새롭게 설계하고 개발하는, 누가 봐도 무모한 방법을 선택했다. 나뿐 아니라 같이 과제에 참여하는 교수님들이나 담당 부서에서도 이 때문에 여러 어려움이 있었을 것은 충분히 이해가 갈 것이다. 그렇지만 사실 처음부터 이렇게 하려고 작정하고 이 과제를 맡았다. 2000년 공중보건의사를 대상으로 시작한 우리나라 역학조사관은 그동안 제도적으로도 많은 변화가 있었고, 사스, 메르스, 신종플루, 코로나19 유행 등을 연속적으로 겪으면서 그 일선에 있는 역학조사관의 역할과 역량이 새롭게 개발되어야 할 시점이라고 판단했기 때문이다.

역학조사관은 전문적인 역학조사와 질병관리를 위해서 특별한 교육과 훈련을 받는다. 역학조사관의 모델이 된 미국의 EIS (Epidemic Intelligence Service)도 의사, 수의사, 보건학 박사 등의 학위나 경력을 가진 사람을 뽑아서 2년간 훈련을 시켜 까다로운 조건을 만족해야 수료를 한다. 이런 전문 역학훈련과정은 일찍부터 여러 나라에서 수용해서 진행해 왔는데, 우리나라는 동남아 국가들보다 늦게 2000년에야 공중보건의사들을 대상으로 역학조사관제도를 시작했다. 이때 선진국의 교육과정을 받아들여 2년 동안 한 번의 기본교육(3~4주)과 연 수회의 지속교육, 학술대회 참여 등의 교육과정을 만들었다. 선진국의 경우 이런 집합교육은 전체 학습내용의 10% 정도를 다루고 실제 90%의 교육은 경험 있는 튜터들로부터 On-the-Job으로 진행된다.

우리나라의 역학조사관 제도는 특히 2015년 메르스, 2020년 코로나-19를 겪으면서 많이 바뀌었다. 역학조사관 수가 크게 확장되었고 이에 따라 역학조사관으로 지원하는 사람들의 배경도 매우 다양해졌다. 이 과정에서 역학조사관 내에서도 교육기간과 수료조건을 달리하는 세분화도 일어났다. 한마디로 우리나라 역학조사관제도는 다른 국가들과는 다른 모습으로 발전해 왔다고 할 수 있다. 현재 역학조사관들의 다양한 학문적 배경, 소속기관에서의 역할과 위치, On-the-Job 훈련이 어려운 조직환경 등 여러 가지가 외국의 기존 EIS 제도와 다르다. 그러면 우리 현실에 맞는 역학조사관의 모습과 교육은 어떤 것이 되어야 할까? 그간 고민해 오던 이 문제에 대해서 여러 사람의 지혜를 모아서 그 첫 대답을 이 책에 담았다.

『역학조사관 교육 표준교재』는 총 3권으로 개발하였다. 분량 때문에 나눈 것은 아니다. 새로운 학습목표가 어느 정도 분명해지자 이 학습목표를 달성하기 위한 학습내용의 성격에 따라서 기초적인 내용

은 1권, 기본적으로 알아야 하지만 활용적인 측면은 2권으로 처음부터 분리해서 기획하였다. 그래서 제1권은 「역학 입문 과정」이라고 부제를 달았고, 제2권은 「역학 역량 강화 과정」이라고 부제를 달았다. 이 1, 2권 모두 기본교육에 활용되는 것으로 개발되었기 때문에 실제 그렇게 활용될 것이다. 그렇지만 1권은 처음 역학조사관 교육을 받는 사람이 많이 활용할 것이고 역학조사관 기본교육을 받았으나 업무에서 어려움을 겪을 때 다시 찾아보는 내용은 2권에 더 많을 것이다.

제3권도 처음부터 기획된 것이다. 3권은 2편으로 나누어져 있는데 1편은 역학조사과정을 실제 상황에 따라 학습할 수 있는 코스워크 또는 문제바탕학습(PBL) 교재로 개발된 9개의 학습모듈로 구성되었다. 이 중 5개는 이번에 새롭게 개발되었고, 4개는 기존에 개발되어 있었던 것을 발굴해 수정보완하여 교재에 넣었다. 이것은 시작이고, 외국처럼 역학조사 학습모듈 교재들이 계속 개발되어 여기에 새로운 유행 모듈이 추가되고 기존 것은 계속 보완과정을 거쳐서 우리나라만의 훌륭한 역학조사 교재로 발전하기를 바라는 마음이다. 3권의 2편은 기존의 역학조사관 선배들이 직접 작성한 역학조사 사례집(2008~2023년)들에서 뽑아서 실었다. 가급적 다양한 감염병과 전파경로, 다양한 대상들을 조사한 것 중에서 선택을 했다. 이것은 이와 유사한 사례를 역학조사할 때 직접 도움이 될 수 있기를 바라는 마음에서 기획한 것이다. 또 각 사례마다 직접 의견을 달았는데, 이것은 각 사례를 평가하기 위한 것이 아니라 무조건 따라하기보다는 비평적으로 사례들을 읽을 수 있도록 하기 위한 목적으로 작성한 것이다.

2024년 7월, 이 과제에 대해서 첫 자문위원회를 했을 때 설명을 들은 자문위원들은 1년 안에 책이 나오는 것은 거의 불가능하다고 솔직한 자문을 해주셨다. 숨차게 달려오느라 사실 세부적인 사항까지 모두 완벽할까를 생각하면 걱정이 앞선다. 그럼에도 일단 누군가는 첫발을 내디뎌야 하기 때문에, 일단 모두 힘을 합쳐서 첫발을 내딛어 본다. 이 교재가 향후 계속 판을 거듭하면서 더욱 발전하기를 바란다.

같이 애써 주신 편집위원분들과 짧은 집필 기간에도 좋은 원고를 보내주신 필자분들, 원고 검토에 참여해 주신 질병관리청 각 부서의 담당자분들의 도움이 없었으면 이 첫발을 내딛는 것조차 불가능했을 것이다. 또 연구실의 장준수, 박하린, 박세열 조교 선생님들도 같이 고생을 했다. 이 모든 분들에게 깊은 감사를 드리는 마음뿐이다.

2025년 5월

편집위원장, 한국역학회 회장

천병철

편집 위원

편집위원장 천병철 고려대학교 의과대학

편집위원 박지혁 동국대학교 의과대학
손현진 동아대학교 의과대학
유석현 가톨릭대학교 의과대학
이주형 전북대학교 의과대학
질병관리역량개발담당관 질병관리청

자문위원 곽 진 전북대학교 의과대학
박수경 서울대학교 의과대학
이종구 서울대학교 의과대학
정해관 성균관대학교 의과대학

집필진 고재영 질병관리청 대변인실
곽 진 전북대학교 의과대학
김종헌 성균관대학교 의과대학
김진용 인천광역시의료원 감염내과
김철우 고려대학교 의과대학
김현정 고려대학교 의과대학
김홍빈 서울대학교 의과대학
남해성 충남대학교 의과대학
류소연 조선대학교 의과대학
문성준 경찰대학교 경찰학과
박지혁 동국대학교 의과대학
손현진 동아대학교 의과대학

우성일 충북대학교 의과대학
유대성 전남대학교 수의과대학
유명순 서울대학교 보건대학원
유석주 동국대학교 의과대학
유석현 가톨릭대학교 의과대학
윤태호 부산대학교 의과대학
이 관 동국대학교 의과대학
이무식 건양대학교 의과대학
이승은 대한감염관리간호사회
이윤재 식품의약품안전처 대변인실
이윤현 남서울대학교 보건행정학과
이재현 전북대학교 의과대학
이주형 전북대학교 의과대학
이철호 질병관리청 대변인실
정지아 한국보건복지인재원
정해관 성균관대학교 의과대학
조민철 고려대학교 의과대학
지희정 충북대학교 정보통계학과
진선미 한국보건복지인재원
천병철 고려대학교 의과대학
최보율 한양대학교 의과대학
최상아 국립중앙의료원
최영준 고려대학교 의과대학
최원석 고려대학교 의과대학
허중연 아주대학교 의과대학

검토진

고재영 질병관리청 대변인실
구현숙 질병관리청 위기관리총괄과
권윤형 질병관리청 질병관리역량개발담당관
김경남 질병관리청 질병관리역량개발담당관
김다슬 질병관리청 감염병관리과
김동근 질병관리청 호흡기감염병TF
김동명 질병관리청 비축물자관리과
김승혁 질병관리청 검역정책과
김아라 질병관리청 호흡기감염병TF
김옥수 질병관리청 검역정책과
김유미(A) 질병관리청 위기관리총괄과
김유미(B) 질병관리청 감염병정책과
김유연 질병관리청 질병관리역량개발담당관
김종희 질병관리청 인수공통감염병관리과
김지아 질병관리청 호흡기감염병TF
김지은 질병관리청 결핵정책과
박영준 질병관리청 결핵정책과
박재선 질병관리청 진단관리총괄과
박진욱 질병관리청 검역정책과
신나리 질병관리청 항생제내성관리과
신인숙 질병관리청 인수공통감염병관리과
양진선 질병관리청 감염병관리과
유 미 질병관리청 질병관리역량개발담당관
유효순 질병관리청 질병감시전략담당관
이데보라 질병관리청 감염병관리과
이상원 질병관리청 질병데이터과학분석관
이선영 질병관리청 인수공통감염병관리과
이소담 질병관리청 인수공통감염병관리과

이신영 질병관리청 항생제내성관리과
이철호 질병관리청 대변인실
이현정 질병관리청 진단관리총괄과
이형민 질병관리청 예방접종정책과
정연화 질병관리청 검역정책과
하 진 질병관리청 감염병정책과

목차

제1장 역학자료 분석 역량강화 ····· 001

제1절 통계학의 기초 개념 ····· 002

1. 역학조사에서 통계학 ····· 002
2. Excel 기초 ····· 009
3. 기술통계: 표와 그래프의 작성 ····· 019
4. 표본분포와 추정 ····· 034
5. 통계적 가설검정 ····· 043
6. 범주형 자료의 비교 ····· 064
7. 상관분석의 이해 및 실습 ····· 076
8. 회귀분석의 이해 및 실습 ····· 079

제2절 보건통계학의 응용 ····· 089

1. 비모수 검정 ····· 089
2. 로지스틱 회귀분석 ····· 096
3. 생존분석 ····· 106

제2장 감시자료의 분석 역량강화 ····· 111

제1절 감시자료의 시계열 추세파악 및 실습 ····· 112

1. 시계열 자료의 이해 ····· 112
2. 시계열 자료로서의 감시자료 분석 ····· 116
3. 감시자료의 시계열 분해 ····· 120
4. 맺음말 ····· 122

제2절 R surveillance 패키지를 이용한 유행 감지방법 ····· 124

1. R surveillance 패키지 ····· 124
2. R surveillance 패키지의 기본 사용법 ····· 125
3. R surveillance 패키지를 이용한 유행 발견의 알고리듬 ····· 126
4. R surveillance 패키지를 이용한 유행 발견 실습 ····· 130
5. 결론 ····· 138

제3장 유행역학조사 역량강화 ····· 141
제1절 유행곡선을 이용한 공통노출일과 평균잠복기 추정법 ····· 142
1. 단일공동노출의 유행곡선 ····· 143
2. 유행곡선에서 공통노출일 계산원리 ····· 144
3. 공통노출일 계산 사례 ····· 146
4. 유행곡선으로 평균 잠복기 계산하기 ····· 147
5. 공통노출일과 평균잠복기 추산결과의 해석 ····· 149
제2절 층화분석과 다변량분석을 이용한 연관성 지표의 보정 ····· 152
1. 연관성 지표와 교란인자 ····· 152
2. 연관성 지표와 효과수정인자 ····· 154
3. 교란인자를 보정한 연관성지표의 계산 ····· 156
4. 맺음말 ····· 163
제3절 유행조사자료를 이용한 감염재생산수 추정법 ····· 164
1. 감염재생산수 ····· 164
2. 감염재생산수의 구성요소 ····· 165
3. 감염재생산수의 추정방법 ····· 166
4. R_0 통계패키지를 이용한 감염재생산수 추정법 ····· 170
5. 맺음말 ····· 174
제4절 수인성 및 식품매개감염병 유행 시 역추적 조사와 환경 평가의 활용 ····· 176
1. 역추적 조사 ····· 177
2. 환경 평가 ····· 179

제4장 감염병진단 역량강화 ····· 183
제1절 감염병 진단의 정확성과 신뢰도에 대한 이해 ····· 184
1. 개요 ····· 184
2. 타당도와 신뢰도의 개념 ····· 185
3. 타당도의 평가 ····· 186
4. 신뢰도의 평가 ····· 192

제2절 차세대염기서열 기반 유전체분석검사의 이해 ····· 195
1. 차세대염기서열분석의 검사원리 ····· 196
2. 역학조사를 위한 전장유전체분석 ····· 200
3. 메타게놈 분석 ····· 203
4. 전장유전체 염기서열분석 데이터를 활용한 역학조사 도구 ····· 206
5. 기타 역학조사를 위한 검사방법들 ····· 208

제5장 예방접종관리 역량강화 ····· 211
제1절 관찰역학연구를 통한 지역사회 백신효과의 측정 ····· 212
1. 백신의 효과 측정 ····· 212
2. 백신효과 측정을 위한 관찰역학 연구 ····· 215
3. 맺음말 ····· 223
제2절 예방접종 의향에 대한 이해와 소통 중심의 대응 방안 ····· 225
1. 배경 ····· 225
2. 예방접종 의향과 행태 이해 ····· 228
3. 백신 주저의 대응 방안 ····· 238

제6장 의료관련감염과 의료관련감염병의 특성과 관리 ····· 247
제1절 의료관련감염과 의료관련감염병의 정의와 특성 ····· 248
1. 의료관련감염의 정의 ····· 248
2. 국내 의료관련감염병의 정의 및 종류 ····· 249
3. 의료관련감염 및 의료관련감염병의 특성 ····· 250
4. 국내 법정 의료관련감염병의 주요 특성 ····· 251
제2절 감염예방 및 관리의 종류와 적용범위 ····· 254
1. 감염예방 및 관리의 정의 및 목적 ····· 254
2. 의료관련감염 및 의료관련감염병의 전파 방식 ····· 254
3. 의료관련감염 및 의료관련감염병 관리의 기본 원리 ····· 255
4. 감염예방 및 관리를 위한 주의지침 및 종류 ····· 256

제3절 의료관련감염병 감시체계 ····· 260
1. 의료관련 감염병 감시 대상 ····· 260
2. 의료관련감염병 감시체계 및 발생현황 ····· 260
제4절 의료관련감염 및 의료관련감염병 유행 시 역학조사 단계와 방법 ····· 266
1. 의료관련감염 및 의료관련감염병 역학조사의 기본 개념 ····· 266
2. 의료관련감염 및 의료관련감염병 유행의 인지와 보고 ····· 269
3. 역학조사의 단계와 방법 ····· 272

제7장 검역 및 해외유입 감염병의 관리 ····· 283
제1절 검역의 역할과 방법 ····· 284
1. 검역의 역할 ····· 284
2. 검역의 기능 ····· 290
3. 검역방법 ····· 296
제2절 해외유입 감염병의 예방과 관리 ····· 303
1. 해외여행과 감염병 ····· 303
2. 해외감염병 감시체계 ····· 305
3. 우리나라 해외유입 감염병 현황 ····· 308
4. 해외유입 감염병의 예방 ····· 311
5. 해외유입 감염병의 관리 ····· 313
6. 맺음말 ····· 314

제8장 원헬스와 기후변화 ····· 317
제1절 원헬스개념과 구성요소 ····· 318
1. 원헬스개념과 구성요소 ····· 318
제2절 인수공통감염병의 원헬스적 접근 ····· 323
1. 인수공통감염병 관리의 원헬스 적용과 활용 ····· 323
2. 인수공통감염병 유행시 동물 및 환경보건 ····· 330

제3절 항생제내성 관리와 원헬스 ····· 340
1. 항생제 ····· 340
2. 항생제내성 ····· 341
3. 항생제내성의 발생과 전파 ····· 343
4. 항생제내성의 중요성 ····· 345
5. 항생제내성을 막기 위한 방안들 ····· 346
제4절 기후변화와 감염병 ····· 351
1. 기후변화와 감염병 ····· 351
2. 매개체 감염병 ····· 353
3. 수인성 식품매개 감염병 ····· 358
4. 호흡기 감염증 ····· 360

제9장 과학적 보고서 작성과 발표 ····· 363
제1절 역학논문 비평적으로 읽기 ····· 364
1. 연구설계의 구분 ····· 364
2. 역학연구의 평가를 위한 주요 고려사항 ····· 367
3. 연구설계별 평가도구 ····· 368
4. 요약 ····· 373
제2절 체계적문헌고찰 논문의 해석 ····· 375
1. 체계적문헌고찰 정의와 특성 ····· 375
2. 체계적문헌고찰의 결과 해석 시 방법론적 고려사항 ····· 376
3. 체계적문헌고찰의 분석결과 해석 ····· 379
4. 요약 ····· 387
제3절 감시분석보고서의 작성 ····· 389
1. 개요 ····· 389
2. 감염병 감시분석보고서의 자료원과 데이터 정리 ····· 390
3. 변수의 속성에 따른 표와 그래프의 작성과 서술 ····· 391
4. 보고서 작성 시 유의 사항과 수정대비표의 작성 ····· 395

5. 맺음말 ····· 396
제4절 역학조사 보고서의 작성 ····· 397
1. 유행 및 역학의 개념 ····· 397
2. 유행 역학조사 단계 ····· 398
3. 역학조사 보고서의 개요 ····· 400
4. 역학조사 보고서 작성방법 ····· 402
제5절 초록 작성 및 학술 발표 ····· 408
1. 초록 작성하기 ····· 408
2. 학술 발표하기 ····· 410
제6절 STROBE와 ORION 가이드라인의 이해:
역학조사 결과의 논문 작성 및 연구결과의 공표 ····· 416
1. 관찰역학연구의 논문작성 가이드: STROBE ····· 417
2. 역학조사의 논문작성 가이드: ORION ····· 419
3. 역학조사결과 해석의 주의점 ····· 422
4. 유행 역학조사 논문에 대한 제언 ····· 423
5. 맺으며 ····· 425

부록: 찾아보기 ····· 427

역학조사관 교육
표준교재

Textbook For
Korea Field Epidemiology
Training Program





1장

역학자료 분석 역량강화







제1절 통계학의 기초 개념
제2절 보건통계학의 응용

집필자 지희정

※엑셀 실습자료: 한국역학회 홈페이지 – 자료실

제1절

통계학의 기초 개념

1 역학조사에서 통계학

통계학(statistics)은 자료를 효과적으로 수집하고 체계적으로 정리하며 분석하는 데 필요한 과학적 이론과 실증적 기법을 다루는 학문이다. 이를 통해 우리는 자료로부터 유의미한 정보를 얻고, 그 결과를 바탕으로 합리적인 의사결정을 내릴 수 있다. 연구 설계, 데이터 수집, 분석, 해석에 이르는 모든 단계에서 중요한 역할을 하며, 과학, 의학, 경제학, 사회학 등 다양한 분야에서 폭넓게 활용된다. 보건역학 연구에서도 통계학은 기본이 되는 필수 도구다.

역학연구는 주로 건강과 질병의 분포 및 결정 요인을 규명하는 것을 목표로 하며, 이를 위해 다양한 데이터를 체계적으로 다루어야 한다. 통계학은 관심 집단의 일부에서 얻은 정보를 바탕으로 전체 모집단에 대한 신뢰할 수 있는 추정을 가능하게 한다. 이를 통해 역학연구는 질병의 위험 요인을 평가하고, 예방 전략을 수립하며, 공중보건 정책을 과학적으로 뒷받침할 수 있다. 예를 들어, "독감 백신 접종이 고령층에서 독감 발병을 줄이는 데 얼마나 효과적인가?"라는 문제를 해결하기 위해 역학조사관은 자료를 수집하고 통계 분석을 실시하여 실제로 고령층에서 독감 발생률이 감소했는지를 확인한다. 통계 분석 결과 고령층에서 독감 발생률이 유의미하게 감소한 것으로 나타난다면, 이를 근거로 지역 보건소와 병원에서 고령층을 대상으로 독감 백신 접종 캠페인을 확대하는 보건 정책을 수립할 수 있다.

1. 기술통계와 추론통계

기술통계(descriptive statistics)와 추론통계(inferential statistics)는 데이터를 분석하고 해석하는 통계학의 두 가지 주요 분야이다. 기술통계는 데이터를 요약하고 정리하여 관심 있는 연구 집단의 특성을 설명하는 방법으로, 데이터의 전반적인 경향을 이해하는 데 사용된다. 반면, 추론통계는 모집단의 특성인 모수를 추정하거나 가설을 검정하는 데 활용되는 방법이다. 분석의 목적과 자료의 형태에 따라 적합한 방법을 선택하며, 보통 기술통계를 통해 데이터의 경향을 파악한 후, 이를 바탕으로 모집단에 대한

추론을 수행하는 접근법이 일반적이다.

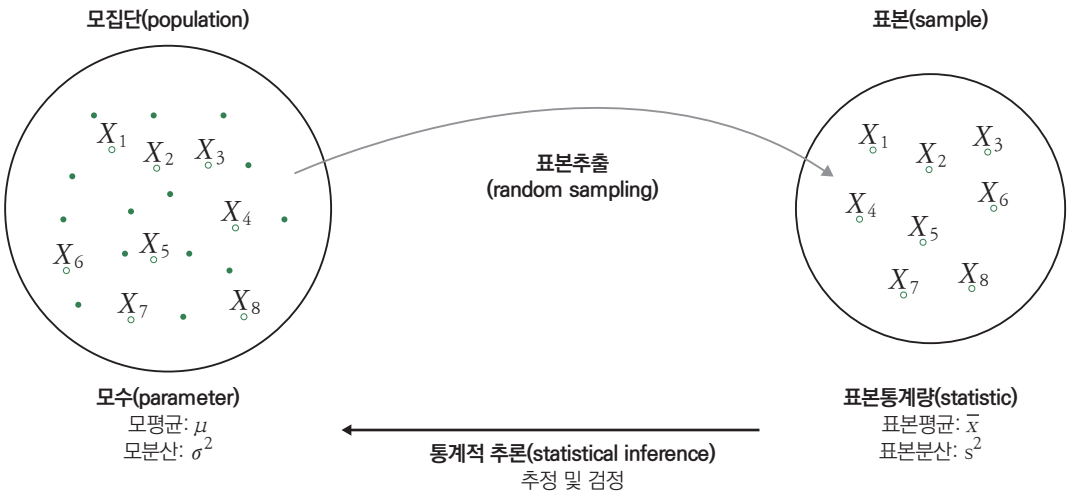


그림 1-1-1. 통계적 추론

기술통계는 많은 양의 데이터를 그림과 몇 개의 수치로 기술하고 요약한다. 예를 들어, 평균, 중앙값, 표준편차와 같은 수치값은 데이터의 전반적인 경향을 이해하는 데 도움을 준다.

추론통계는 관심 연구집단인 모집단(population)의 일부인 표본(sample) 데이터를 바탕으로 모집단에 대한 특성을 추론하는 것이 목적으로, 표본에서 얻어진 통계량(statistic)을 이용하여 모집단의 특성인 모수(parameter)를 추정(estimation)하거나 가설검정(hypothesis testing)을 하는 방법이다. 모수는 평균 μ, 표준편차 σ, 모비율 π, 상관계수 ρ와 같이 주로 그리스-로마 문자로 표현하며, 통계량은 표본평균 $\bar{x}$, 표본 표준편차 s, 표본비율 p, 표본 상관계수 r과 같이 주로 영문자로 표현한다.

2. 모집단과 표본

모집단(population)은 연구자가 관심을 가지고 조사하고자 하는 대상의 전체 집합을 의미한다. 이는 특정 특성이나 조건을 공유하는 사람, 사물, 사건 등을 모두 포함하며, 이론적으로 무한하거나 유한한 크기를 가질 수 있다. 모집단의 크기와 구성은 연구의 목적과 범위에 따라 결정된다. 예를 들어, "대한민국의 고령자(65세 이상 인구)의 독감 백신 접종률"을 알고자 한다면, 여기서 모집단은 대한민국의 모든 고령자(65세 이상 인구)이다.

표본(sample)은 모집단의 일부로, 모집단에서 추출된 하위 집단을 의미한다. 모집단의 크기가 크거

나 무한한 경우, 모집단을 적절히 대표하는 표본을 추출하여 자료를 수집한다. 이후 표본에서 얻어진 통계량(statistics)을 이용해 모집단의 특성인 모수(parameter)를 추정하거나, 모수에 대한 가설검정을 수행할 수 있다. 예를 들어, "대한민국의 고령자(65세 이상 인구)의 독감 백신 접종률"을 알고자 한다면, 대한민국의 모든 고령자의 독감 백신 접종 여부 자료를 수집하는 데 시간과 비용 등 현실적 제약이 있을 수 있다. 따라서 대한민국의 모든 고령자를 적절히 대표할 수 있는 고령자 1,000명을 추출하여 설문조사를 통해 독감 백신 접종 여부 자료를 얻는 것이 효율적이다. 이때 뽑힌 고령자 1,000명이 표본이다.

표본은 모집단 전체가 아닌 일부를 추출한 것이기 때문에, 표본 추출 과정에서 발생하는 변동이 존재한다. 이로 인해 표본에서 얻은 평균이나 분산은 모집단의 평균이나 분산과 차이가 있을 수 있다. 이러한 차이를 표본오차(sampling error)라고 한다. 또한, 자료를 수집, 분석, 결과를 기록하는 과정에서 발생할 수 있는 모든 오류를 비표본오차(non-sampling error)라고 한다. 비표본오차에는 설문 문항의 오류, 데이터 입력 실수, 응답자의 부정확한 답변, 무응답 등이 포함된다. 표본에서 얻은 결과를 모집단에 일반화하기 위해서는 표본이 모집단을 적절히 대표할 수 있도록 추출하는 것이 중요하다. 적절한 표본 추출은 표본오차를 최소화하고 연구 결과의 신뢰성과 타당성을 높인다.

3. 자료의 종류

데이터는 관찰이나 측정을 통해 얻어진 사실이나 값의 집합을 의미한다. 데이터는 숫자, 문자, 기호 등의 형태로 나타날 수 있으며, 변수와 데이터를 이해하는 것은 통계 분석의 출발점이 된다.

변수는 연구나 분석에서 관찰하거나 측정할 수 있는 속성이나 특성을 의미한다. 변수가 없다면 데이터를 수집하거나 분석할 수 없으므로, 변수는 통계학에서 가장 기본적인 개념 중 하나이다. 변수는 변화할 수 있는 값을 가지며, 이러한 변화는 연구자가 연구 질문을 통해 밝혀내고자 하는 주요 요소가 된다. 예를 들어, 사람의 나이, 성별, 키, 체중, 혈압 등은 모두 변수가 될 수 있다. 데이터의 형태는 다양하기 때문에, 적절한 통계 분석 방법을 선택하려면 각 변수가 어떤 형태의 데이터를 가지는지 파악하는 것이 중요하다.

1) 측정 척도에 따른 분류

자료는 측정 척도에 따라 네 가지로 구분할 수 있다. 이 구분은 데이터를 분석하는 방법과 도구를 선택할 때 중요한 기준이 된다.

(1) 명목형

범주 간의 순서가 없으며 단순히 구분하기 위해 사용된다. 명목형 자료에서 숫자는 단순히 코드로 사용될 뿐 실제 연산에는 의미가 없다. 예를 들어, 성별(남성, 여성)이나 혈액형(A, B, AB, O)이 이에 해당한다.

(2) 서열형(Ordinal)

범주 간에 순서가 있지만, 그 간격은 일정하지 않다. 만족도(매우 불만족, 불만족, 보통, 만족, 매우 만족) 같은 변수가 이에 해당한다. 예를 들어, 만족도 조사에서 "보통"과 "만족" 간의 간격이 "매우 불만족"과 "불만족" 간의 간격과 동일하지 않을 수 있다.

(3) 등간형(Interval)

범주 간의 순서와 간격이 일정하지만, 절대적인 "0"의 기준이 없다. 예를 들어, 섭씨 온도나 IQ 점수가 있다. 섭씨 온도에서 0도는 절대적인 "없음"을 의미하지 않고 단순히 특정한 기준점일 뿐이다. 이러한 특성 때문에 비율 계산은 의미가 없다.

(4) 비율형(Ratio)

절대적인 "0"이 존재하며, 비율 비교가 가능하다. 키, 몸무게, 수입 등이 비율형 변수의 예이다. 예를 들어, "몸무게가 0 kg"이라는 것은 아무것도 없음을 의미하며, 40 kg은 20 kg의 두 배라는 비율 계산이 가능하다.

변수의 측정 척도를 이해하는 것은 매우 중요하다. 이는 측정 척도에 따라 적용할 수 있는 통계 분석 방법이 달라지기 때문이다. 측정 척도는 필요에 따라 변환될 수 있으며, 이를 통해 변수의 활용 범위를 확장하거나 분석의 목적에 맞게 조정할 수 있다. 예를 들어, 등간형 변수인 연령(0세에서 100세)을 서열형 변수로 변환하는 경우를 생각해보자. 연령 데이터를 "연령대"로 나누어 구간화할 수 있다. 이렇게 변환하면 연령이라는 연속적인 값을 서열형 변수로 바꿀 수 있어, 연령대별 분포를 파악하거나 그룹 간 차이를 분석하는 데 활용할 수 있다.

이와 같이, 변수의 측정 척도를 적절히 변환하고 활용하면 분석 목표에 맞는 데이터를 생성하고 보다 효율적으로 통계 분석을 수행할 수 있다. 하지만 모든 변수들이 서로 변환이 가능한 것은 아니다. 명목형 변수는 서열형 변수, 등간형 변수, 비율형 변수로 변환이 불가능하다. 반대로 비율형 변수는 등간형 변수, 서열형 변수, 명목형 변수로 변환이 가능하다. 즉, 변수에도 계층 구조가 있는데 비율형 변수가 가장 높고 명목형 변수는 가장 낮으며, 그림 1-1-2와 같이 높은 구조에서 낮은 구조로만 변환이 가

능하다.

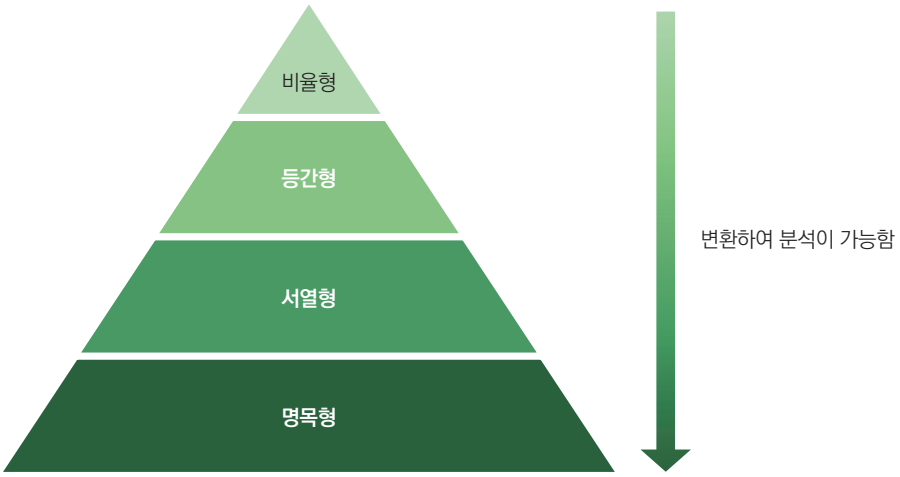


그림 1-1-2. 측정 척도의 계급

자료 변환의 예

ID	성별	연령	접종여부
1	Female	12	Yes
2	Male	23	No
3	Male	45	No
4	Female	37	Yes
⋮	⋮	⋮	⋮
100	Male	65	Yes

범주형 자료로 변환

ID	성별	연령	접종여부
1	Female	청소년	Yes
2	Male	청년	No
3	Female	중년	No
4	Male	청년	Yes
⋮	⋮	⋮	⋮
100	Male	노년	Yes

변수의 구분

구분	설명	예시
명목형(Norminal)	• 범주 간 상호 배타적이며, 순서가 없다.	성별, 혈액형, 지역
순서형(Ordinal)	• 범주 간 상호 배타적이며, 순서가 있으나 간격은 일정하지 않다.	사회경제적 수준, 학력수준
등간형(Interval)	• 값들이 동일한 간격으로 배열되었다.	온도, 날짜
	• 절대적인 0점이 없다.	
	• 덧셈, 뺄셈은 가능하지만 곱셈과 나눗셈은 의미가 없다.	
비율형(Ratio)	• 값들이 동일한 간격으로 배열되었다.	키, 몸무게, 나이
	• 절대적인 0점이 존재한다.	
	• 덧셈, 뺄셈, 곱셈, 나눗셈이 모두 의미가 있다.	

2) 역할에 따른 변수의 분류

변수는 연구 설계에서 그 역할에 따라 독립변수와 종속변수로 나뉜다. 독립변수는 종속변수에 영향을 주는 변수이며, 종속변수는 연구자의 관심변수이자 독립변수에 의해 영향을 받는 변수이다. 보통 독립변수를 X, 종속변수를 Y로 표현한다. 종속변수에 영향을 미치는 독립변수가 여러 개인 경우(X_1, X_2, …)도 있고, 종속변수가 여러 개인 경우(Y_1, Y_2, …)도 있다.

(1) 독립변수

종속변수에 영향을 미치는 변수로 종속변수를 설명해주는 변수이다. 연구자가 조작하거나 통제하여 종속변수에 미치는 영향을 확인하려는 대상이다. 예를 들어, 흡연 여부는 폐암 발생 여부에 영향을 줄 수 있는 독립변수(independent variables)다. 독립변수는 실험 설계에서 필수적으로 고려해야 할 요소다. 독립변수는 다른 용어로 설명변수(explanatory variable), 실험변수(experimental variable), 또는 처치변수(treatment variable)이라고도 한다.

(2) 종속변수

독립변수의 영향을 받는 변수로 독립변수의 조작에 의존한다. 위 예에서 폐암 발생 여부가 종속변수(dependent variables)에 해당한다. 종속변수는 연구에서 결과로 나타나는 변수이며, 독립변수와의 관계를 규명하는 데 초점이 맞춰진다. 종속변수는 반응변수(response variable) 또는 결과변수(outcome variable)이라고도 한다.

3) 자료의 특성에 따른 분류

통계 분석에서는 보통 자료를 양적 자료(quantitative data)와 질적 자료(qualitative data)로 구분한다. 양적 자료는 수치형 자료(numerical data)라고도 불리며, 수치형 값들로 이루어진 자료이다. 정수값을 가지는 이산형 자료와 연속적인 수치값을 갖는 연속형 자료가 이에 해당한다. 반면, 질적 자료는 취할 수 있는 값이 서로 구분 가능한 범주로 표현되는 자료이며, 명목형 자료와 순서형 자료가 이에 속한다. 질적 자료는 다른 용어로 범주형 자료(categorical data)라고 한다. 이 구분은 데이터의 표현 방식과 분석 방법을 결정하는 데 중요한 역할을 한다.

4) 시간적 특성에 따른 분류

데이터는 관측 시점이나 기간에 따라 다음과 같이 구분된다. 시간적 특성을 이해하면 데이터의 구조를 명확히 파악하고, 적절한 분석 기법을 선택할 수 있다.

(1) 단면 데이터

특정 시점에서 관측된 데이터를 의미한다. 단일 시점의 데이터를 사용하여 특정 시점의 상태를 분석하거나 비교할 수 있다. 예를 들어, 2024년 한 해 동안 조사된 인구 통계는 단면 데이터(cross-sectional data)다. 단면 데이터는 주로 비교 연구나 현황 분석에 사용된다.

(2) 시계열 데이터

시간의 흐름에 따라 관측된 데이터를 말한다. 이러한 데이터는 시간에 따른 변화나 패턴을 분석하는 데 유용하다. 예를 들어, 2010년부터 2024년까지의 연도별 경제 성장률이 시계열 데이터(time-series data)다. 시계열 데이터는 예측 모형 개발이나 추세 분석에 활용된다.

(3) 패널 데이터

동일한 대상을 여러 시점에 걸쳐 관측한 데이터를 의미한다. 예를 들어, 한 개인의 건강 상태를 10년간 추적하여 관찰한 데이터가 이에 해당한다. 패널 데이터(panel data)는 개인 간 차이와 시간에 따른 변화를 동시에 분석할 수 있어, 동적인 변화를 분석하거나 인과관계를 분석할 때 특히 유용하다.

4. 자료의 형태에 따른 대략적인 분석 방법

다음 표는 자료의 형태에 따라 자주 사용되는 통계분석 방법들을 정리한 표이다. 연구 설계에서 변수

는 그 역할에 따라 독립변수와 종속변수로 구분할 수 있다. 독립변수는 종속변수에 영향을 미치는 변수이며, 독립변수에 따라 변하는 결과를 종속변수라 하였다. 물론 표에서 정리한 분석 방법외에도 분석의 목적, 상황에 따라 다양한 분석 방법이 적용될 수 있다.

자료의 형태에 따른 분석 방법

종속변수의 형태	독립변수의 개수	독립변수의 형태	분석방법
연속형	1개	이분형	독립 이표본 t-검정
			대응 표본 t-검정
		다항형	분산분석
		연속형	상관분석
			회귀분석
	2개 이상	연속형, 범주형	회귀분석
이분형	1개	이분형	카이제곱 검정
		다항형	카이제곱 검정
이분형	1개	연속형	로지스틱 회귀분석
	2개 이상	연속형, 범주형	로지스틱 회귀분석

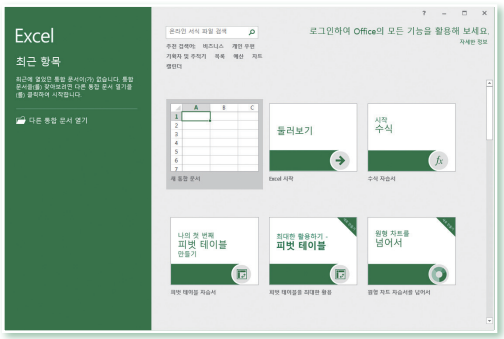
2 Excel 기초

Excel은 Microsoft에서 개발한 스프레드 시트(spread sheet) 프로그램으로, 행렬 형태로 데이터를 관리하는 데 적합하다. 데이터는 효율적인 처리를 위해 일반적으로 행렬 형태로 구조화된다.

Excel의 스프레드시트는 1행부터 시작하여 최대 1,048,576개의 행으로 구성되며, 열은 A, B, C, …, XFD까지 총 16,384개로 구성된다. Excel은 데이터를 다양한 형식으로 저장하고 관리할 수 있을 뿐만 아니라, 수식과 함수를 활용해 복잡한 계산을 자동화할 수 있다. 이 외에도 데이터 분석, 계산, 그래프 작성, 피벗 테이블 생성 등의 다양한 기능을 제공한다.

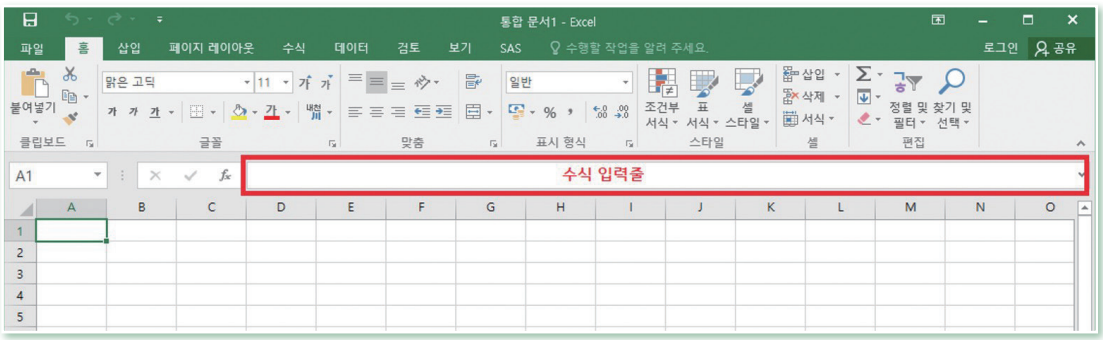
1. Excel의 시작

아래의 그림은 Excel을 처음 실행했을 때의 화면이다. '새 통합 문서'를 선택하여 시작할 수 있다.

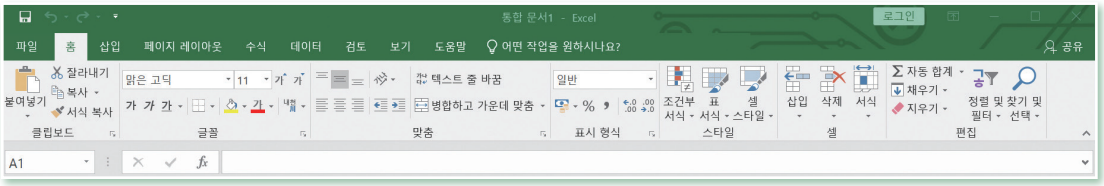


2. Excel의 구성

Excel 화면의 구성을 살펴보면, 다양한 작업을 효율적으로 수행할 수 있도록 설계되어 있음을 알 수 있다. 먼저, 행과 열로 이루어진 각각의 직사각형 단위를 셀(Cell)이라고 하며, 이는 데이터를 입력하고 계산할 수 있는 기본 단위이다. 각 셀은 고유한 주소(예: A1, B2, C3)를 가지며, 주소는 열 문자를 먼저 쓰고 행 번호를 뒤에 적는 방식으로 나타낸다.



셀을 클릭하면 해당 셀이 활성화 상태가 되며, 활성화된 셀은 주위에 사각형 테두리가 표시된다. 예를 들어, 위 그림에서 활성화된 셀의 주소는 A1이다. 활성화된 셀에 데이터를 입력하거나 수정할 때는 셀 내부뿐만 아니라 상단의 수식 입력줄에서도 실시간으로 입력 내용을 확인할 수 있다. 이러한 셀들로 이루어진 작업 공간을 워크시트(worksheet)라고 한다. 워크시트는 데이터를 입력하고 처리할 수 있는 주요 작업 공간이며, 위 그림에서는 Sheet 1이 활성화되어 있다. 새로운 워크시트를 추가하려면 화면 하단에 있는 + 버튼을 클릭하면 된다. 여러 워크시트를 하나로 묶으면 통합 문서를 구성할 수 있으며, 통합 문서 하나는 하나의 파일로 저장된다. 통합 문서는 여러 워크시트를 포함하므로 데이터를 체계적으로 정리하고 관리할 수 있는 구조를 제공한다.



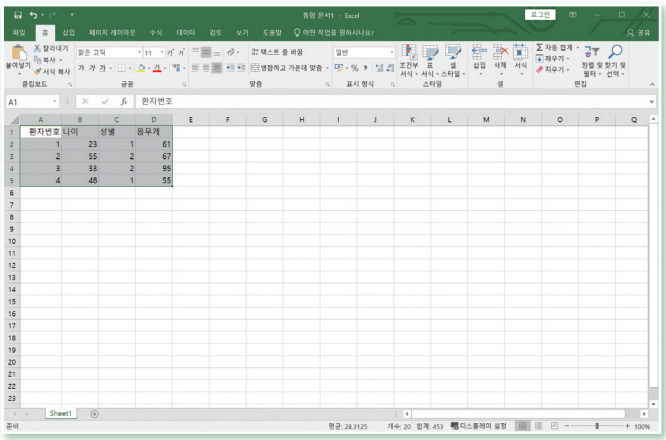
다음으로 Excel의 메뉴 화면을 살펴보자. Excel 메뉴는 파일, 홈, 삽입, 페이지 레이아웃 등 여러 탭으로 구성되어 있으며, 각 탭은 관련된 기능들을 묶어 제공한다. 각 탭에 포함된 버튼은 사용자가 필요한 기능을 빠르게 실행할 수 있도록 설계되어 있다. 버튼을 클릭하면 해당 기능이 즉시 실행되며, 버튼 오른쪽 아래에 있는 화살표 아이콘을 클릭하면 더 구체적인 설정을 변경할 수 있다.

이제 Excel에서 데이터를 입력하고 저장하는 실습을 진행해 보자. 다음과 같이 데이터를 입력해 보고 저장하는 방법을 학습해 보자.

	A	B	C	D
1	환자번호	나이	성별	몸무게
2	1	23	1	61
3	2	55	2	67
4	3	33	2	95
5	4	48	1	55

① 먼저, A1셀을 클릭하여 활성화한다. A1셀에 “환자번호”를 입력하고 키보드의 Tab 키를 누르면 B1셀로 이동된다. B1셀에 “나이”를 입력하고 Enter 키를 누르면 B2셀로 이동된다. 위와 같은 방식으로 각 셀에 데이터를 입력한다.

② 이번에는 데이터 값들이 셀의 우측에 위치하도록 변경하여 보자. 먼저, 변경하고자 하는 셀들을 선택해야 한다. 여러 개의 셀을 한 번에 변경하기 위해 마우스 드래그 또는 Shift 키와 방향키를 사용하여 셀의 범위를 지정해준다. 본 실습에서는 A1을 클릭한 후, D5까지 마우스를 끌면 된다.



③ 셀의 범위를 지정해준 상태에서 메뉴의 홈 탭의 맞춤 그룹으로 이동한다. 오른쪽 맞춤 단추를 클

릭하여 데이터가 셀의 오른쪽으로 정렬되도록 한다. 정렬을 마친 문서는 다음과 같다.

	A	B	C	D
1	환자번호	나이	성별	몸무게
2	1	23	1	61
3	2	55	2	67
4	3	33	2	95
5	4	48	1	55

④ 상단 메뉴에서 파일 탭을 클릭한 후 다른 이름으로 저장 메뉴를 클릭한다.

⑤ 저장할 위치를 선택한다(예: 바탕화면). 여기서는 파일 이름을 "환자데이터"라고 하겠다. 파일 형식은 기본값인 Excel 통합 문서로 설정하고 저장 버튼을 클릭한다. 이때 기본 형식으로 저장하면 파일 확장자는 xlsx로 지정된다. 저장 버튼을 클릭하여 파일을 저장하면, 이 문서는 "환자데이터.xlsx"라는 이름으로 저장된다.

『Tip 빠른 저장하기』

작업 중에는 Ctrl + S 단축키를 사용하여 저장을 빠르게 할 수 있다.

3. Excel 함수

Excel에서 유용한 기능 중 하나는 다양한 수식과 함수를 사용하여 복잡한 계산을 간단히 수행할 수 있다는 점이다. 함수는 함수 이름, 인수, 괄호로 구성되며, 항상 등호(=)로 시작한다. 등호는 수식의 시작을 나타내며, 등호를 쓰지 않고 함수 이름을 바로 쓰는 경우 Excel은 이를 함수가 아닌 텍스트로 인식한다.

함수 이름은 함수의 종류를 나타낸다. 예를 들어, SUM, AVERAGE, IF 등의 함수가 있다.

인수는 함수가 작업을 수행하는데 필요한 입력값이다. 함수의 인수 자체가 함수인 경우에 "중첩"되었다고 하며, 7단계까지 함수를 중첩할 수 있다.

괄호는 함수의 이름 뒤에 열리며, 인수를 묶는 역할을 한다.

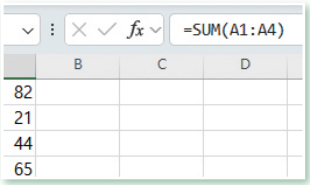
- 함수의 형식: =함수이름(인수, 인수, …, 인수)
- 예제: =AVERAGE (A1:A8, SUM(1,2,3))

특히, 통계 분석이나 데이터 관리에서 자주 사용하는 함수들을 실습을 통해 학습해보자.

1) SUM 함수

SUM 함수는 지정된 범위 내 숫자들의 합을 계산해주는 함수이다. 데이터를 집계하거나 합산해야 할 때 자주 사용된다.

[실습] A1, A2, A3, A4의 합을 A5 셀에 구해보자.

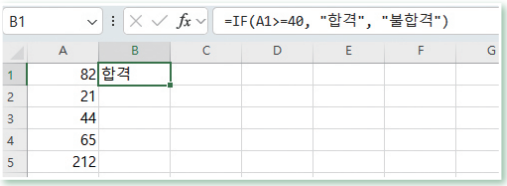


- 방법 1: A5 셀을 클릭한 후, "=A1+A2+A3+A4"를 입력한다.
- 방법 2: A5 셀을 클릭한 후, "=SUM(A1:A4)"를 입력한다.

2) IF 함수

IF 함수는 조건에 따라 다른 값을 반환하는 함수이다. 데이터를 조건에 따라 결과를 출력하거나 분류하는 작업에 사용된다. 특히, 연속형 데이터를 특정 구간으로 나누어 범주형 데이터로 분류하거나, 기존 범주형 데이터를 몇 개의 범주로 합쳐 새로운 재범주화 변수를 생성하는 데 유용하다.

[실습] B1 셀에 A1 셀의 값이 40 이상이면 "합격", 그렇지 않으면 "불합격"을 반환해보자.

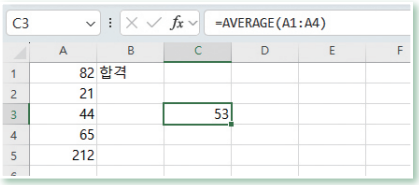


B1 셀을 클릭한 후, "=IF(A1 >=40, "합격", "불합격")"을 입력한다.

3) AVERAGE 함수

AVERAGE 함수는 지정된 범위 내 값들의 평균(산술평균)을 계산하는 함수이다. 평균은 데이터 값의 합을 해당 데이터 값의 개수로 나눈 값으로, 데이터의 중심 경향을 파악할 때 자주 사용된다.

[실습] A1 셀부터 A4 셀까지의 평균을 계산하여 C3 셀에 구해보자.

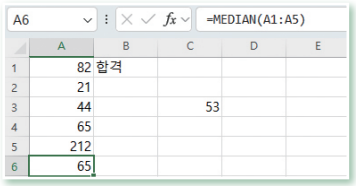


C3 셀을 클릭한 후, "=AVERAGE(A1:A4)"을 입력한다.

4) MEDIAN 함수

MEDIAN 함수는 지정된 범위 내 값들의 중위수를 계산하는 함수이다. 중위수는 데이터를 크기 순으로 정렬했을 때 가운데에 위치한 값을 의미하며, 데이터의 중심 경향을 파악할 때 자주 사용된다.

[실습] A1 셀부터 A5 셀까지의 값에서 중앙값을 계산하여 A6 셀에 구해보자.

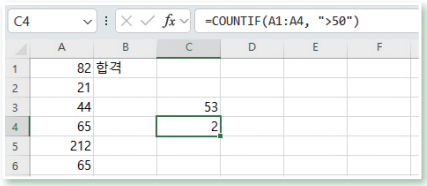


A6 셀을 클릭한 후, “=MEDIAN(A1:A5)”을 입력한다.

5) COUNTIF 함수

COUNTIF 함수는 지정된 범위에서 특정 조건에 맞는 셀의 개수를 계산하는 함수이다. 데이터를 조건에 따라 필터링하거나, 특정 값이나 범위에 해당하는 데이터 개수를 파악할 때 유용하다.

[실습] A1 셀부터 A4 셀에서 50보다 큰 값의 개수를 세어보자.



C4 셀을 클릭한 후, “=COUNTIF(A1:A4, “>50”)”을 입력한다.

4. Excel의 참조: 절대참조와 상대참조

Excel에서 참조란 셀의 위치를 표시하는 방법으로, 수식에서 특정 셀이나 셀 범위를 지정하여 해당 데이터를 활용할 수 있도록 하는 기능이다. 참조는 상대참조, 절대참조, 혼합참조 세 가지 유형이 있다.

1) 상대참조

‘A1’과 같은 형태로 셀 위치를 작성하며, 셀을 복사하거나 이동할 때 참조하는 셀의 위치가 상대적으로 변경된다. 예를 들어, ‘A1+A2’를 A3에 입력한 후 이를 복사하여 B3에 붙여넣으면, 수식은 자동으로 ‘B1+B2’로 변경된다.

2) 절대참조

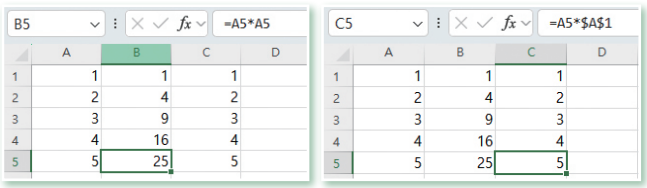
'$A$1'과 같은 형태로 셀의 위치를 고정하여 참조하는 방식이다. $기호를 사용하여 행과 열을 고정하는 것이다. 예를 들어, '$A$1+$A$2'를 A3에 입력한 후 이를 복사하여 B3에 붙여넣으면, 수식은 여전히 '$A$1+$A$2'로 고정된다.

3) 혼합참조

'A$1'과 같은 형태로 참조하는 셀의 행 또는 열 중 하나만 고정하는 방식이다. 예를 들어, 'A$1+A$2'를 A3에 입력한 후 이를 복사하여 B3에 붙여넣으면, 수식은 여전히 'B$1+B$2'로 행은 변경되고 열은 유지된다.

셀 참조를 이용하면 하나의 함수식에서 흩어져 있는 여러 개의 데이터를 동시에 처리하여 작업효율을 높이는데 도움을 준다. 실습을 통해 절대참조와 상대참조의 차이를 확인해보자.

① A1부터 A5까지의 셀에 각각 1, 2, 3, 4, 5를 입력한다.
② B1 셀에 "=A1*A1"를 입력하고 Enter를 누른다.
③ B1 셀을 복사하여 B2부터 B5까지 붙여넣기 한다.
④ C1 셀에 "=A1*$A$1"을 입력하고 Enter를 누른다.
⑤ C1 셀을 복사하여 C2부터 C5까지 붙여넣기 한다.
⑥ B5 셀과 C5 셀의 결과값을 확인해보자.



이때, B5 셀의 결과값은 "=A5*A5"로 변경되어야 하며, C5의 셀의 결과값은 "=A5*$A$1"이 되어야 한다.

이렇게 상대참조는 셀 주소가 수식을 복사하거나 이동할 때 상대적으로 변경되고, 절대참조는 셀 주소가 고정되어 수식을 복사하거나 이동해도 참조 셀 값이 유지된다.

5. 데이터 입력 규칙

통계 분석을 위해 데이터를 입력할 때는 몇 가지 규칙을 준수해야 한다. 대부분의 통계 분석 프로그

램은 이러한 입력 규칙을 따르기 때문에, Excel에서 데이터를 입력한 후 다른 프로그램(예: R, SAS, SPSS)으로 데이터를 불러와 분석하려면 이 규칙을 미리 적용해 두는 것이 좋다. 이를 통해 다른 프로그램에서 데이터를 불러올 때 발생할 수 있는 문제를 방지할 수 있으므로, 데이터 입력 시 반드시 이를 염두에 두고 작업하는 것이 바람직하다.

Excel을 사용하면 다양한 통계 분석이 가능하지만, 이 책의 후반부에서 다루는 로지스틱 회귀분석과 생존분석은 Excel에서 기본적으로 제공하지 않는다. 로지스틱 회귀분석과 생존분석을 수행하려면 Excel에서 정리한 데이터를 R, SAS, SPSS와 같은 통계 분석 프로그램으로 불러와 분석해야 한다. 따라서 이러한 분석을 계획하고 있다면, 데이터 입력 규칙을 미리 준수하여 데이터를 정리해 두는 것이 작업 시간을 단축하고 분석 과정에서의 어려움을 줄이는 데 도움이 된다.

통계분석을 위한 데이터 입력 규칙:

<table>
<tr><th>항목</th><th>규칙</th><th>예</th></tr>
<tr><td>변수 이름</td><td>• 첫 번째 행에 변수 이름을 입력
• 영문 사용을 권장
• 알파벳, 숫자, 밑줄(_)만 사용
• 첫 글자는 문자로 시작
• 공백 및 특수문자 미사용</td><td>age
gender
income_2024</td></tr>
<tr><td>데이터 정렬</td><td>• 각 열은 하나의 변수를 나타내고 각 행은 하나의 관측값을 나타냄
• 시계열 자료의 경우, 시간이 변함에 따라 행을 구분하여 자료를 입력함</td><td><table><tr><th>ID</th><th>age</th><th>gender</th></tr><tr><td>1</td><td>20</td><td>M</td></tr><tr><td>2</td><td>40</td><td>F</td></tr></table>
시계열 자료
<table><tr><th>time</th><th>incidence</th></tr><tr><td>2024-01-01</td><td>46</td></tr><tr><td>2024-01-02</td><td>72</td></tr><tr><td>2024-01-03</td><td>85</td></tr></table></td></tr>
<tr><td>데이터 입력</td><td>• 수치형 변수에는 숫자만 입력함
• 문자형 변수는 동일한 형식의 텍스트를 사용
• (예: 성별 변수에 M,F 라고 입력하다가 중간에 Male, Female로 입력하면 다른 값으로 인식)</td><td><table><tr><th>ID</th><th>age</th><th>gender</th></tr><tr><td>1</td><td>20</td><td>M</td></tr><tr><td>2</td><td>40</td><td>F</td></tr></table></td></tr>
</table>

① 첫 번째 행에 아래의 변수를 작성한다. 변수명은 영어를 사용한다.

- 입력할 변수명: ID, age, height, sex

② 각 변수에 대해 하나의 행이 한 대상자, 하나의 열이 대상자의 특정 속성을 나타내도록 아래의 예시 데이터를 작성한다.

	A	B	C	D
1	ID	age	height	sex
2	1	23	166	F
3	2	43	182	M

이렇게 보통 하나의 행은 대상자를 의미하며, 하나의 열은 대상자의 특정 속성을 나타내는 변수로 구분한다. 첫 번째 행은 각 변수의 이름을 적으며, 변수 이름은 가능한 영문으로 작성하는 것이 좋다. 숫자도 포함될 수 있으나, 되도록 변수 이름의 첫 글자는 영문자로 시작하는 것이 좋다.

6. 기존 변수를 활용해 새로운 변수 만들기

ex_survey_data.xlsx 데이터는 다양한 건강 관련 변수들을 포함하고 있다. 그중, T_AGE는 나이를 의미하는 변수이다. 이번에는 T_AGE(나이) 변수를 활용하여 나이대 변수를 생성하는 실습을 진행하여 보자. T_AGE 변수를 기준으로 나이대를 다음과 같이 분류한다:

- 0~29세: 청년
- 30~59세: 중년
- 60세 이상: 노년

① 엑셀의 새로운 열에 AGE_GROUP이라는 이름을 추가한다.

② 아래와 같은 IF 함수를 사용하여 값을 계산한다.

```
=IF(C2<=29, "청년", IF(C2<=59, "중년", "노년"))
```

③ 첫 번째 셀에 수식을 입력한 후, 나머지 셀로 드래그하여 모든 셀에 적용한다.

	A	B	C	D
1	T_ID	T_SEX	T_AGE	AGE_GROUP
2	K_BASE_03	2	53	중년
3	K_BASE_04	2	62	노년
4	K_BASE_00	1	40	중년
5	K_BASE_04	2	51	중년
6	K_BASE_02	2	40	중년
7	K_BASE_01	1	58	중년
8	K_BASE_03	2	44	중년
9	K_BASE_06	2	45	중년
10	K_BASE_07	2	46	중년
11	K_BASE_05	2	62	노년
12	K_BASE_08	2	53	중년
13	K_BASE_00	2	49	중년
14	K_BASE_08	2	52	중년
15	K_BASE_01	1	68	노년
16	K_BASE_00	2	56	중년
17	K_BASE_06	2	53	중년
18	K_BASE_04	1	40	중년
19	K_BASE_08	1	42	중년
20	K_BASE_07	1	60	노년
21	K_BASE_05	1	49	중년

Sheet1

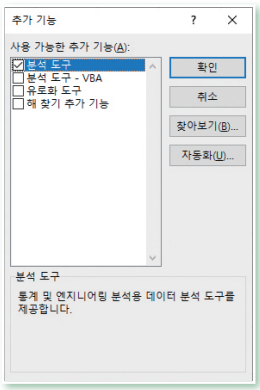
7. 데이터 분석 도구

Excel에서 데이터를 분석하기 위해 데이터 분석 도구를 활성화해야 한다.

① Excel의 상단 메뉴의 파일 → 옵션을 클릭한다.

② 추가기능 탭에서 Excel 추가 기능을 선택하고 하단의 이동 버튼을 클릭한다.

③ 분석 도구를 체크한 뒤 확인 버튼을 클릭한다.



④ 상단의 데이터 Tab에 데이터 분석 메뉴가 추가된 것을 확인할 수 있다.



3 기술통계: 표와 그래프의 작성

데이터 분석의 첫 번째 단계는 데이터를 직관적으로 이해하고 그 특성을 파악하는 것이다. 이를 위해 방대한 측정값들을 시각화하거나 요약 통계를 계산하여 데이터의 전반적인 구조와 특징을 탐구한다. 이러한 탐색적 과정은 데이터를 깊이 이해하고 분석의 방향성을 설정하며, 분석 과정에서 발생할 수 있는 문제를 사전에 확인하는 데 도움을 준다.

탐색적 자료분석은 기술통계 기법을 활용하여 데이터를 요약하고, 변수의 분포를 파악함으로써 데이터의 전반적인 특성을 이해하는 데 초점을 맞춘다. 또한 데이터를 시각적으로 표현하여 변수의 분포, 그룹 간 차이, 변수 간 상관관계 등 데이터의 경향성을 확인하며, 이를 통해 분석의 방향성을 구체화한다. 이 장에서는 데이터를 기술하는 데 사용되는 그래프와 요약 통계량에 대해 다룬다.

1. 자료의 요약

먼저, 데이터를 몇 가지 기술통계량(descriptive statistics)으로 요약하여 데이터의 분포를 전반적으로 파악하고자 한다. 기술통계는 자료의 유형에 따라 적합한 분석 방법이 다르다. 연속형 자료는 평균과 표준편차와 같은 수치를 사용하여 데이터를 요약하며, 범주형 자료는 빈도와 비율을 통해 자료를 설명한다. 순서형 자료는 최빈값이나 중앙값으로 중심을 표현하고, 범위로 산포도를 나타낼 수 있다. 이처럼 자료의 형태에 따라 적합한 기술통계량은 달라질 수 있다. 다음 절에서는 중심위치를 나타내는 중심척도와 자료의 퍼짐 정도를 설명하는 산포도에 관한 기술통계량을 자세히 소개한다.

1) 범주형 자료의 요약: 빈도

범주형 자료를 기술하기 위해 주로 사용되는 방법은 빈도분석이다. 빈도분석은 각 범주에 속하는 데이터의 빈도(frequency)와 백분율(percentage)을 계산하여 자료의 분포를 요약한다.

- 빈도: 각 범주에 속하는 관측값의 개수
- 백분율: 각 범주의 빈도를 전체 자료의 수로 나누어 백분율로 표현한 값

예를 들어, 연구대상자 1,486명을 대상으로 성별 분포를 파악하고자 한다고 가정해보자. 성별은 “남자(1)” 또는 “여자(2)”로 구분되는 명목형 변수이다. 이 데이터를 바탕으로 빈도분석을 수행하면, 다음

과 같은 결과를 얻을 수 있다:

	빈도	백분율
남자	554	37.3%
여자	932	68.7%

이 분석 결과에서 여자의 수가 남자보다 378명 더 많다는 것을 알 수 있다. 이를 통해 전체 연구대상자 중에서 여자의 비율이 더 높다는 사실을 쉽게 파악할 수 있다. 이러한 빈도분석은 데이터의 분포를 직관적으로 이해하고, 자료의 특성을 명확하게 설명하는 데 유용하다.

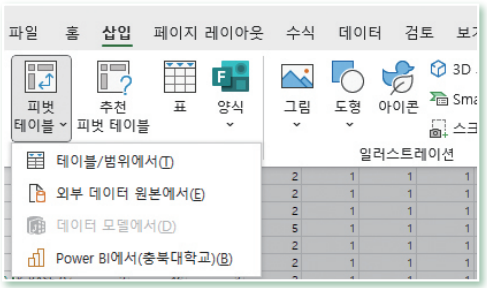
[실습] 빈도분석: 피벗테이블

Excel을 이용하여 실습을 진행해보자. 실습 데이터 ex_survey_data.xlsx는 보건의료연구자원정보센터(Korea Biobank Resource Center, KBRC)에서 제공하는 한국인유전체역학조사사업(Korea Genome and Epidemiology Study, KoGES)의 코호트 자료를 기반으로 생성된 교육용 데이터 중 일부를 추출하여 구성되었다. 이 데이터는 300명의 표본을 대상으로 성별, 만 나이, 월평균 수입을 포함한 기본정보와 질병의 진단여부, 생활습관, 임상검사 결과 지표를 수집한 자료이다. 실습을 진행하기 전, 자세한 변수 목록과 각 변수에 대한 설명은 코드북을 참고하기 바란다.

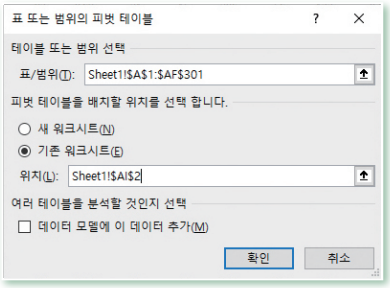
ex_survey_data.xlsx에서 성별에 따른 월평균 수입의 빈도를 분석하기 위해 피벗 테이블을 생성한다.

① 데이터의 아무 값이나 선택한 후, 단축키 Ctrl+A를 누르면 전체 범위가 선택된다.

② 전체 범위를 선택한 후, [삽입] 탭에서 [피벗테이블] 중 테이블/범위에서(T)를 선택한다.

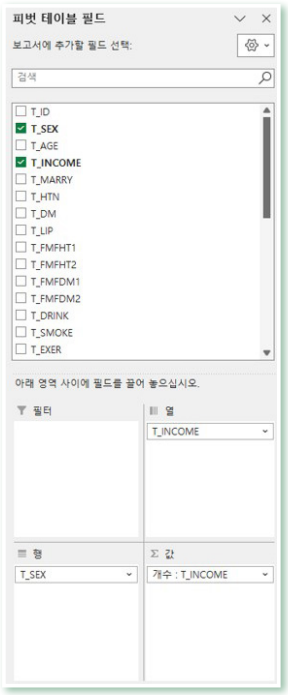


③ [표/범위]는 전체 데이터셋이 되고, 피벗테이블을 배치할 위치를 선택한다. 이는 [기존 워크시트]의 빈 셀을 선택한 후, 확인을 눌러준다.



④ 피벗 테이블 필드 선택하기

성별에 따른 월 평균 수입에 맞는 피벗 테이블 필드를 생성한다. 행에는 성별, 열에는 월 평균 수입, 값에는 월 평균 수입 변수를 드래그하여 각 영역에 변수명을 넣어준다. 값 영역에서 월 평균 수입 변수를 오른쪽 클릭하여 “값 필드 설정”을 누른 후 개수로 변경한다.



⑤ 피벗 테이블이 생성된다.

개수 : T_INCOME	열 레이블								
행 레이블	1	2	3	4	5	6	7	8	총합계
1	4	6	10	14	20	26	12	14	106
2	13	18	17	25	43	42	23	13	194
총합계	**17**	**24**	**27**	**39**	**63**	**68**	**35**	**27**	**300**

2) 연속형 자료의 요약: 중심척도

연속형 측정값들의 분포를 요약하려면 대부분의 자료값이 분포하는 중심에 대한 정보와 그 중심으로부터 얼마나 퍼져 있는지를 설명해야 한다. 이를 통해 연속형 자료의 분포를 대략적으로 파악할 수 있다. 중심을 설명하는 값을 중심척도라고 하며, 평균, 중위수, 최빈값 등이 이에 해당한다. 또한, 퍼짐 정도를 나타내는 산포도에는 분산, 표준편차, 범위, 사분위수 범위 등이 포함된다. 이제 자주 사용되는 중심척도에 대해 알아보자.

(1) 평균

보통 평균(mean)은 산술평균을 의미한다. 산술평균은 모든 자료값을 합한 후 자료의 개수로 나눈 값이다. 산술평균은 자료값을 모두 사용하며, 수학적으로 규명된 성질이 많아 다루기 편하다. 또한, 평균의 표본분포 형태도 잘 알려져 있다. 그러나 이상치(outlier)에 영향을 받기 쉬워, 한쪽으로 치우친 분포에서는 중심에 대한 왜곡된 정보를 제공할 수 있으므로 주의해야 한다.

$$\overline{X}=\frac{x_1+x_2+\cdots+x_n}{n}=\frac{\sum_{i=1}^{n}X_i}{n}$$

(2) 중위수

중위수(median)는 자료를 작은 값부터 큰 값까지 크기순으로 나열했을 때 중앙에 위치한 값을 의미한다. 자료의 개수가 홀수이면 중위수를 비교적 쉽게 찾을 수 있지만, 짝수인 경우에는 정확한 중위수가 존재하지 않는다. 이때는 크기순으로 정렬된 가운데 두 값의 평균을 중위수로 사용한다. 중위수는 자료값 자체보다 순서 정보를 이용해 구하기 때문에 이상치의 영향을 거의 받지 않으며, 분포가 치우친 경우에도 크게 영향을 받지 않는다. 그러나 자료의 전체적인 정보를 충분히 반영하지 못하며, 표본분포가 복잡하다는 단점이 있다.

(3) 최빈값

자료 내에서 가장 빈도가 큰 값을 최빈값(mode)이라고 한다. 보통 범주형 또는 순서형 자료에서 사용하며, 최빈값은 하나 이상일 수도 있고 존재하지 않을 수도 있다.

3) 연속형 자료의 요약: 산포도

산포도는 자료가 흩어진 정도, 즉 변동을 의미한다. 분산(variance), 표준편차(standard deviation),

범위(range), 사분위수 범위(interquartile range) 등이 있다.

(1) 분산

분산(variance)은 각 자료값에서 평균을 뺀 편차의 제곱을 모두 더한 후 (자료의 개수-1)로 나눈 값을 말한다. 분산에 대한 공식은 다음과 같다.

$$s^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

여기서 s^2은 자료의 개수 n이 아니라 $n-1$로 나누는 이유는 엄밀히 말해 n-1로 나누어야 모집단 분산의 불편추정량이 되기 때문이다. 불편추정량은 통계학에서 보다 좋은 추정량으로 간주된다. 분산이 크면 자료가 더 널리 퍼져 있고, 작으면 덜 퍼져 있음을 의미한다. 그러나 원자료의 값을 제곱하여 계산하기 때문에 분산의 단위도 제곱이 되어 직접적인 해석이 어렵다. 분산은 자료값을 모두 사용하여 계산하며, 수리적으로 정의된 공식이 존재한다. 하지만 이상치에 민감하여, 분포가 치우친 경우에는 퍼짐 정도에 대한 왜곡된 정보를 제공할 수 있으므로 주의해야 한다.

(2) 표준편차

표준편차(standard deviation)는 분산의 제곱근(square root) 값이다. 표준편차의 단위는 원자료의 단위와 같아서 변동에 대한 해석을 원자료의 단위에서 해석할 수 있다. 표준편차의 장점과 단점은 분산의 장단점과 동일하다.

$$s=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$$

(3) 범위

범위(range)는 자료의 최대값과 최소값의 차이를 말한다. 계산이 쉬운 장점은 있으나, 이상치가 있는 경우 범위는 실제 전반적인 자료의 변동보다 더 클 수 있다. 또한 자료의 크기가 커질수록 범위는 증가하는 경향이 있다.

(4) 사분위수 범위

사분위수 범위(interquartile range, IQR)는 데이터를 크기 순으로 정렬했을 때 제3사분위수(Q3)와 제1사분위수(Q1)의 차이를 의미한다.

- 제1사분위수(Q1): 데이터의 하위 25% 지점에 해당하는 값.
- 제3사분위수(Q3): 데이터의 하위 75% 지점에 해당하는 값.

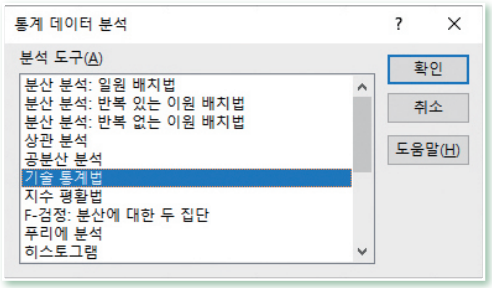
사분위수 범위는 데이터의 중앙 50% 범위를 나타내며, 양 극단 50%에 해당하는 값들은 제외하므로 이상치에 덜 민감하다. 따라서 데이터의 분포가 한쪽으로 치우친 경우에도 변동성을 효과적으로 측정하는 데 유용하다. 그러나 단 두 개의 관측값(Q1, Q3)만 사용하여 계산되므로 데이터 전체의 변동성을 완전히 반영하지 못할 수 있으며, 수리적인 공식이 정의되지 않았다는 단점이 있다.

4) 데이터 분석 도구: 기술 통계법

연속형 변수인 T_AGE(만 나이)의 기술통계량을 구해보자.

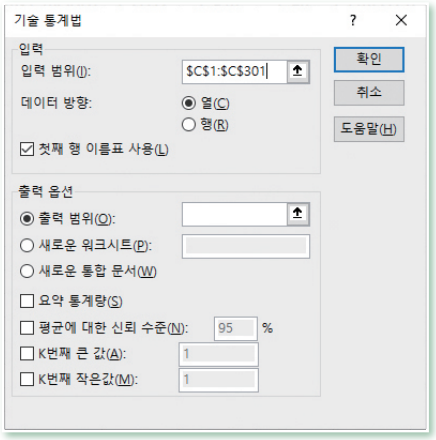
① [데이터] 탭 [분석] 그룹에서 데이터분석을 누른다.

② 분석 도구 중 기술 통계법을 선택한 후, 확인을 누른다.

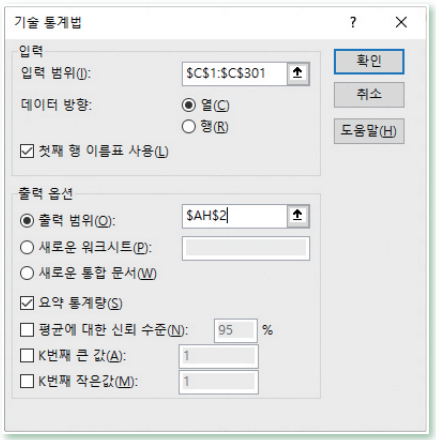


③ 입력 범위를 설정한 후, 데이터 방향을 선택한다.

만약, 입력 범위(I)에 필드명도 포함한 경우, '첫째 행 이름표 사용(L)'을 선택한다.



④ 빈 셀 하나를 출력 범위로 설정을 한 후, '요약 통계량'을 선택한다.



⑤ 결과는 선택된 셀로부터 오른쪽으로 결과가 출력된다.

T_AGE	
평균	53.23333
표준 오차	0.477214
중앙값	53
최빈값	40
표준 편차	8.265589
분산	68.31996
첨도	-0.95461
왜도	0.086279
범위	30
최소값	40
최대값	70
합	15970
관측수	300

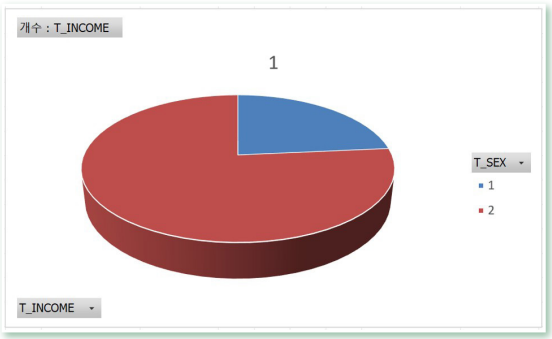
2. 자료를 그래프로 표현하기

1) 범주형 자료: 파이차트

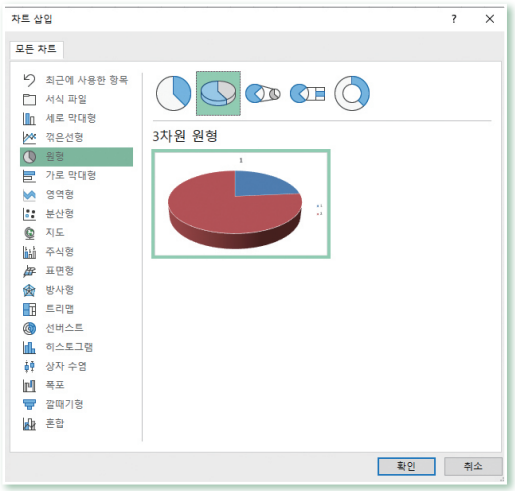
파이차트(pie chart)는 하나의 원을 여러 조각으로 나눈 그래프이다. 각 조각의 크기는 범주형 변수의 각 범주에 대한 비율이나 관측값의 개수를 나타낸다. 주로 전체 대비 각 범주의 구성 비율을 시각적으로 표현할 때 사용된다. 그러나 범주의 개수가 많을 경우, 범주 간 구성 비율이 유사하면 차이를 명확하게 구분하기 어려울 수 있다. 이러한 경우 파이차트는 범주 간의 구성 비율 차이를 효과적으로 반영하기 어렵다.

[실습] Excel을 이용하여 성별을 파이차트로 나타내고자 한다.

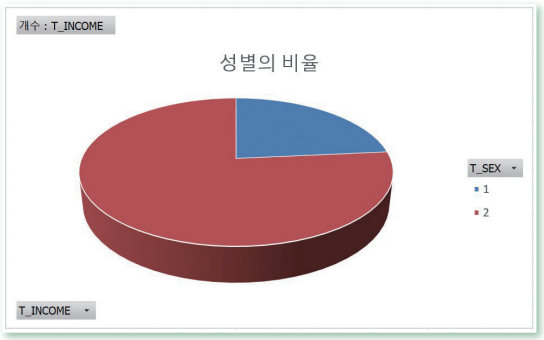
① 앞의 예제에서 생성한 피벗 테이블을 이용하여 성별을 파이차트로 시각화하고자 한다.



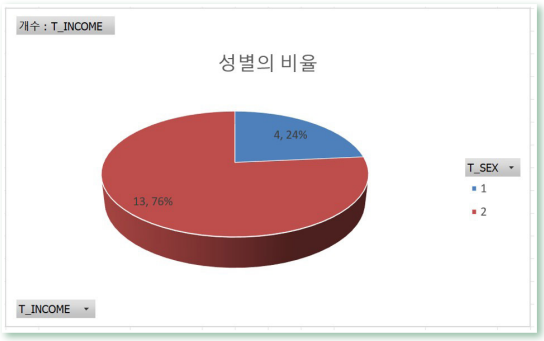
② 피벗테이블을 선택한 후, [피벗 테이블 분석] 탭에서 [피벗차트]을 선택하여 원하는 원형그래프를 선택한다.



③ 원형그래프에 제목을 작성하고, 데이터 레이블을 추가한다.



④ 값 (4,13)을 선택한 후, 마우스 오른쪽을 클릭하여 [데이터 레이블 서식]에서 레이블 내용을 변경해 주었다. 항목이름과 백분율이 나타나도록 했다.



2) 범주형 자료: 막대그래프

막대그래프(bar chart)는 자료를 범주별로 구분하여 수평 또는 수직 막대로 표현한 그래프이다. 각 막대의 길이는 해당 범주의 빈도에 비례한다. 범주 간 빈도가 유사할 경우, 파이차트보다 막대그래프가 범주 간 차이를 더 명확하게 나타내는 데 적절하다.

3) 연속형 자료: 히스토그램

히스토그램(histogram)은 연속형 자료의 분포를 시각적으로 나타내는 그래프이다. 자료를 일정한 구간(bins)으로 나누고, 각 구간에 속하는 자료의 개수를 막대의 높이로 표현한다. 히스토그램은 막대도표와 유사하지만, 연속형 자료의 빈도를 나타내는 그래프이므로 연속형 자료의 특성을 반영하여 막대 간 간격이 없이 표현된다는 특징이 있다. 또한, 구간을 어떻게 나누느냐에 따라 분포 모양이 달라 보일 수 있으므로, 적절한 구간 크기를 선택하는 것이 중요하다. 역학조사의 유행곡선은 시간이 연속형이므로 히스토그램으로 그린다.

[실습] 연속형 변수인 "만 나이"의 히스토그램을 그려보자.

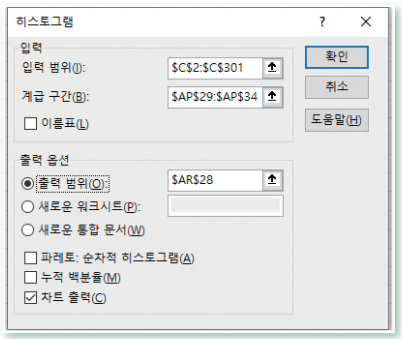
① [데이터분석] 도구를 이용하여 히스토그램을 그리기 위해서는 별도의 계급 구간 데이터를 생성해야한다. 만 나이는 40세부터 70세까지로 구성되어 있다. 계급 구간을 다음과 같이 정의하면 45세 미만, 45세 이상~50세 미만, … , 65세 이상~71세 미만으로 설정할 수 있다.

계급구간
45
50
55
60
65
71

② [데이터] 탭 [분석] 그룹에서 [데이터분석]을 누른다. 분석 도구 중 히스토그램을 선택한 후, 확인을 누른다.

③ 입력 범위에는 만_나이 열을 지정하고, 계급 구간에는 사전에 준비된 계급 구간 열을 지정한다. 출력 범위에는 히스토그램 값이 생성될 빈 셀을 선택한다.

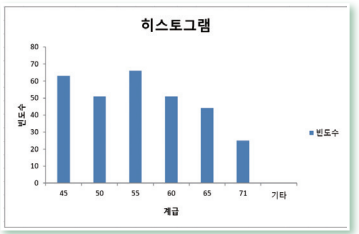
계급구간
45
50
55
60
65
71



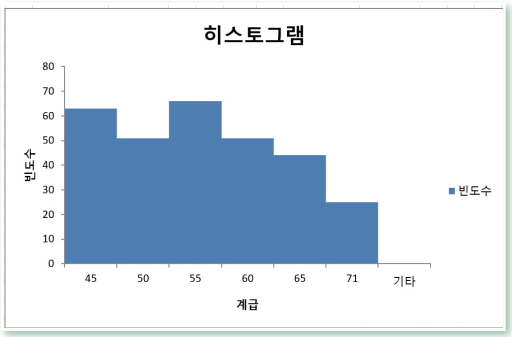
단, 히스토그램을 그리는 경우, 입력 범위(I)는 필드명을 제외한 값부터 지정한다.
빈 셀 하나를 출력 범위로 설정을 한 후, 원하는 차트 옵션을 선택한다.

④ 출력 범위로 선택된 빈 셀부터 오른쪽으로 결과가 출력된다.

계급	빈도수
45	63
50	51
55	66
60	51
65	44
71	25
기타	0



⑤ 축 설정과 간격 너비를 설정한다. 축을 선택한 후 마우스 오른쪽 버튼을 클릭하고, [축 서식]을 선택한다. y축은 0~80으로 지정한다. 히스토그램에서 막대 간격을 없애려면, 막대를 선택한 뒤 [데이터 계열 서식]에서 간격 너비를 0%로 설정한다.



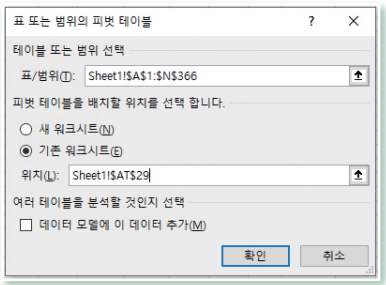
4) 시간에 따라 측정된 자료: 선그래프

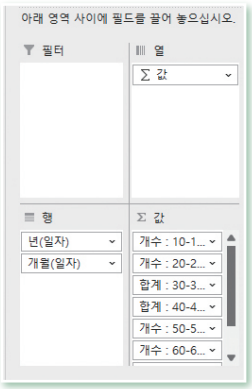
선그래프는 시간에 따른 연속형 변수의 변화를 시각적으로 나타내는 그래프이다. X축에는 시간을, Y축에는 연속형 자료값을 표시하며, 각 점을 선으로 연결하여 변화를 표현한다. 역학 연구에서 활용 예시로는, 감시자료에서 시간에 따른 감염자 수의 변화를 선그래프로 시각화하여 감염병의 유행을 파악하는 것이 있다.

[실습] 코로나바이러스감염증-19 발생과 사망 현황 데이터

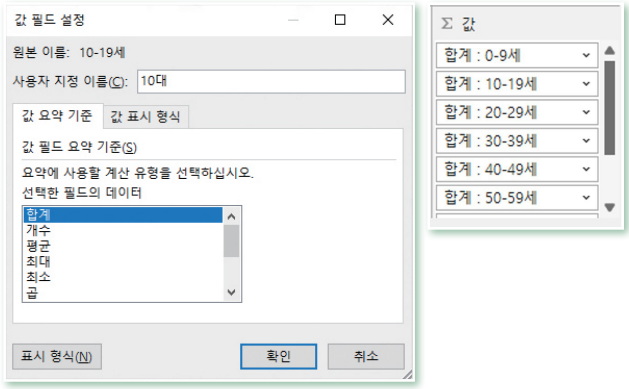
일자별 코로나바이러스감염증-19 발생과 사망 현황 데이터(ex_COVID19_data.xlsx)를 사용하여 선 그래프를 그려보자. "ex_COVID19_data.xlsx" 데이터는 2020년 7월부터 2021년 6월까지 일자별 한국의 Covid-19 신규발생과 사망에 대한 데이터이다.

① 일자별 연령대에 따른 환자 수의 추이를 보고자 한다. 피벗테이블을 활용하여 월별 환자 수를 구한 후 선그래프로 나타내 보자. 데이터를 전체 선택 후, [삽입] 탭에서 [피벗테이블]을 클릭한다. 행에는 년(일자), 개월(일자)를 순서대로 나열하고, 값에는 연령대를 순서대로 나열한다.

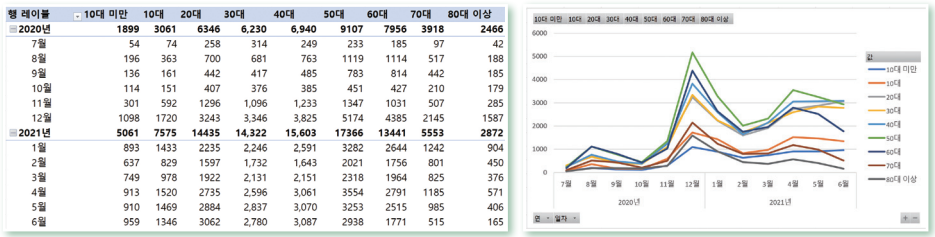




② 값에 있는 연령대를 하나씩 선택하여 오른쪽 클릭 후 [값 필드 설정]에서 값 필드 요약 기준을 “합계”로 변경해준 후, 사용자 지정 이름을 간략하게 수정한다. 모든 연령대에 이 방법을 적용하여 피벗 테이블을 완성한다.



③ 완성된 피벗 테이블을 전체 선택한 후, [피벗 테이블 분석] 탭에서 [피벗 차트] 중 꺾은선형을 선택하여 원하는 선그래프를 완성한다.



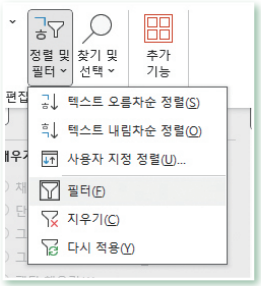
행 레이블	10대 미만	10대	20대	30대	40대	50대	60대	70대	80대 이상
2020년	1899	3061	6346	6,230	6,940	9107	7956	3918	2466
7월	54	74	258	314	249	233	185	97	42
8월	196	363	700	681	763	1119	1114	517	188
9월	136	161	442	417	485	783	814	442	185
10월	114	151	407	376	385	451	427	210	179
11월	301	592	1296	1,096	1,233	1347	1031	507	285
12월	1098	1720	3243	3,346	3,825	5174	4385	2145	1587
2021년	5061	7575	14435	14,322	15,603	17366	13441	5553	2872
1월	893	1433	2235	2,246	2,591	3282	2644	1242	904
2월	637	829	1597	1,732	1,643	2021	1756	801	450
3월	749	978	1922	2,131	2,151	2318	1964	825	376
4월	913	1520	2735	2,596	3,061	3554	2791	1185	571
5월	910	1469	2884	2,837	3,070	3253	2515	985	406
6월	959	1346	3062	2,780	3,087	2938	1771	515	165

5) 두 연속형 자료: 산점도

산점도는 일반적으로 두 연속형 변수 간의 관계를 시각화하는 데에 사용된다. 한 변수를 X축에, 다른 변수를 Y축에 배치하여 두 변수간의 관계를 시각적으로 표현한다. 특히, 회귀분석에서는 산점도를 활용하여 X축에 독립변수, Y축에 반응변수를 배치해 두 변수 간의 관계를 나타낸다. 이를 통해 변수 간 상관관계의 경향을 파악하고, 이상치를 확인하는 데 활용할 수 있다.

[실습] 만 나이가 40인 대상자들의 수축기혈압의 평균과 BMI의 관계를 보고자 한다(ex_BMI_data.xlsx).

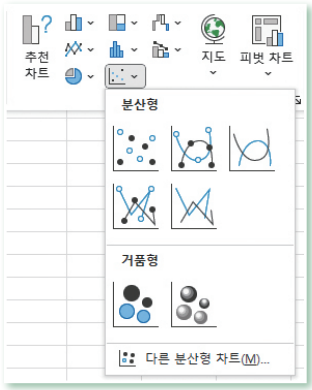
① 나이 변수를 선택한 후, [홈] 탭 [정렬 및 필터]에서 필터를 선택한다.



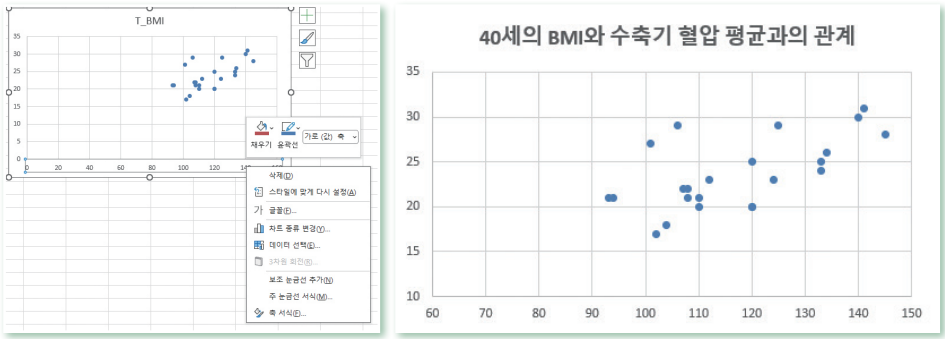
② 만 나이가 40인 대상자를 필터링한 후, 수축기혈압 평균과 BMI 값을 선택하여 새 워크시트에 데이터셋을 생성한다.

	T_SBP	T_BMI
1	141	31
2	124	23
3	110	21
4	140	30
5	133	24
6	101	27
7	93	21
8	145	28
9	108	21
10	110	20
11	120	20
12	107	22
13	112	23
14	104	18
15	120	25
16	94	21
17	125	29
18	108	22
19	106	29
20	133	25
21	102	17
22	120	20
23	134	26

③ 두 변수를 선택하여 [삽입]탭 [차트]에서 원하는 산점도를 선택한다.



④ x축과 y축을 선택한 후 마우스 오른쪽 버튼을 클릭하고, [축 서식]을 선택한다. x축은 최소값은 60, 최대값은 150으로 설정하고, y축은 최소값은 10, 최대값은 35로 변경할 수 있다.

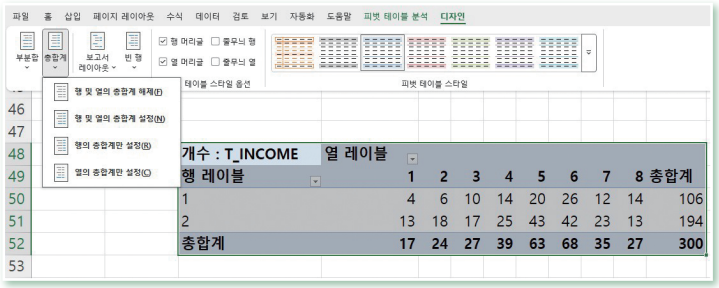


6) 그룹형 막대그래프

이제 Excel이 지원하는 다양한 그래프에 대해 알아보자. 특히, 집단 간 분포를 비교하기 위해 여러 데이터 시리즈를 하나의 그래프에서 표현할 수 있다. 예를 들어, 월소득 집단 간 성별의 분포를 시각적으로 표현하려면 성별의 빈도를 막대그래프로 나타내고, 월소득 범주에 따라 색을 구분하여 나타낼 수 있다. 이를 통해 각 월소득 집단에서 성별 분포의 차이를 쉽게 비교할 수 있다.

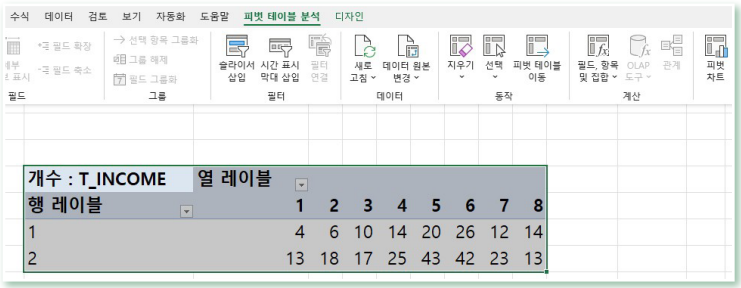
[실습] 월소득 집단 간 성별의 분포 비교

① 이전에 만든 피벗 테이블을 선택하여, [디자인]탭의 [총합계] 중 행 및 열의 총합계 해제를 선택한다. 총합계는 피벗 차트를 그리는 데에는 필요가 없으므로 제외한다.



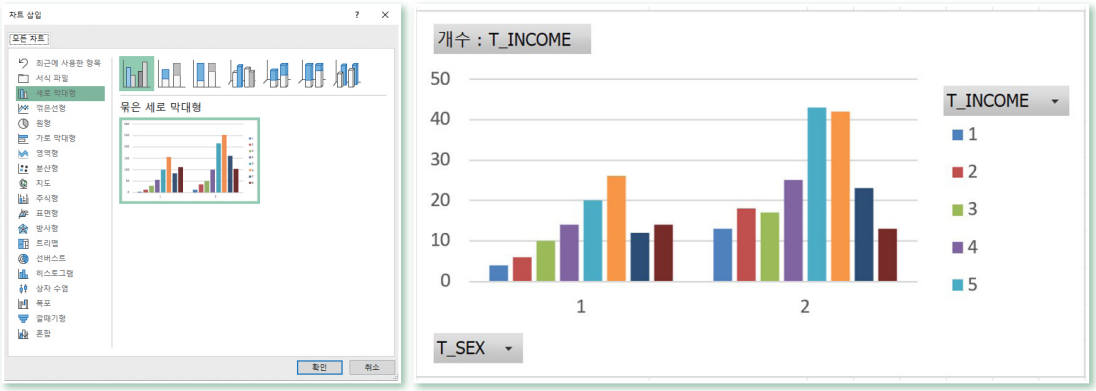
개수 : T_INCOME	열 레이블								
행 레이블	1	2	3	4	5	6	7	8	총합계
1	4	6	10	14	20	26	12	14	106
2	13	18	17	25	43	42	23	13	194
총합계	17	24	27	39	63	68	35	27	300

② 축소된 피벗테이블을 선택한 후, [피벗 테이블 분석] 탭에서 [피벗차트] 선택한다.



개수 : T_INCOME	열 레이블							
행 레이블	1	2	3	4	5	6	7	8
1	4	6	10	14	20	26	12	14
2	13	18	17	25	43	42	23	13

③ 원하는 차트를 선택한다.



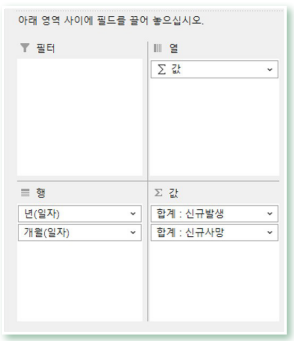
7) 혼합형 차트: 막대그래프와 선그래프

혼합형 차트는 두 개 이상의 서로 다른 그래프 유형을 하나의 차트에 결합하여 자료의 다양한 측면을 동시에 비교하는 방법이다. 역학 연구에서 활용되는 예시로는 질병 발생과 사망을 동시에 비교하거나 예방접종률과 감염률을 동시에 비교할 때가 있다.

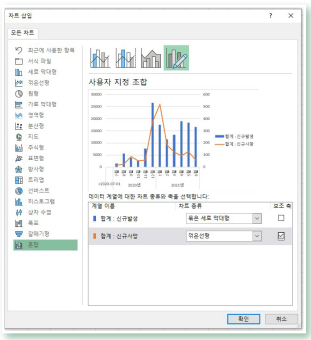
[실습] 월별 Covid-19 신규발생과 사망 그래프

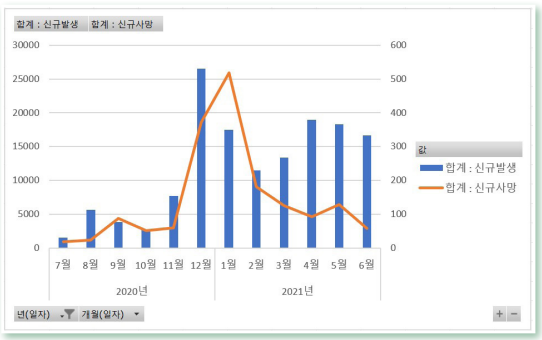
"ex_COVID19_data.xlsx" 데이터를 이용하여 막대그래프와 선그래프의 혼합형 차트를 생성해보자. "ex_COVID19_data.xlsx" 데이터는 2020년 7월부터 2021년 6월까지 일자별 한국의 Covid-19 신규발생과 사망에 대한 데이터이다. 월별 신규발생과 사망의 추이를 확인하고자 한다.

① 일자별 COVID19 데이터에서 피벗테이블 기능을 활용하여 월별 신규발생과 사망자수 피벗테이블을 생성한다. 행에는 년(일자)와 개월(일자)를 순서대로 나열하고, 값에는 신규발생과 신규사망의 합계를 드래그하여 놓는다.



② 피벗 테이블을 전체 선택한 후, [피벗 테이블 분석] 탭에서 [피벗 차트] 중 혼합을 선택하여 원하는 혼합형 차트를 완성시킨다.





4 표본분포와 추정

1. 확률과 확률변수

확률(probability)은 특정 사건이 발생할 가능성을 수량화한 값으로, 0에서 1 사이의 값을 가진다. 확률이 0이면 해당 사건이 발생하지 않음을 의미하며, 확률이 1이면 반드시 발생함을 의미한다.

확률변수(random variable)는 확률적 실험의 결과를 실수로 표현하는 함수이다. 즉, 가능한 모든 실험 결과의 집합인 표본공간(sample space)의 원소를 실수로 대응시키는 함수이다. 확률변수는 이산형 확률변수(discrete random variable)와 연속형 확률변수(continuous random variable)로 나눌 수 있다.

이산형 확률변수는 가능한 값이 유한 개이거나 셀 수 있는 무한 개인 확률변수를 의미한다. 예를 들면, 동전을 10번 던지는 실험을 생각해 보자. 이때 동전을 던져서 나올 수 있는 가능한 결과는 앞면과 뒷면이며, 10번 실험을 했을 때 앞면이 나오는 수를 확률변수 X라고 하자. 확률변수 X가 가질 수 있는 값은 0부터 10까지의 유한 개이므로, 이 경우 확률변수 X는 이산형 확률변수라고 한다. 역학 연구에서 자주 등장하는 이산형 확률변수의 예로는 환자의 감염 여부, 병원에서 하루 동안 보고된 신종 감염병 확진자 수, 특정 질병으로 병원에 입원한 환자의 수 등이 있다. 연속형 확률변수는 무한한 연속적인 값을 가질 수 있는 확률변수이다. 키, 몸무게, 시간 등 연속적인 실수 값을 갖는 변수는 연속형 확률변수에 해당한다.

확률분포(probability distribution)는 확률변수의 부분집합에 확률이 어떻게 할당되는지를 나타내는 것이다. 확률변수의 성질에 따라 이산형 확률분포와 연속형 확률분포로 나뉜다. 이산형 확률분포(discrete probability distribution)는 확률변수가 취할 수 있는 값이 셀 수 있는 경우, 즉 이산적인 경우에 해당한다. 이 경우, 각 가능한 값에 대해 확률을 직접 할당할 수 있으며 이를 나타낸 함수를 확률질량함수(probability mass function)라고 한다. 모든 값에 대한 확률의 합은 1이다. 이산형 확률분포의 대표적인 예로는 베르누이 분포(bernoulli distribution), 이항 분포(binomial distribution), 포아송 분포(poisson distribution) 등이 있다. 연속형 확률분포(continuous probability distribution)는 확률변수가 연속적인 값을 취하는 경우를 의미하며, 개별 값의 확률은 0이고 특정 구간에 속할 확률은 확률밀도함수(probability density function)를 적분하여 계산할 수 있다. 확률밀도함수를 전체 구간에 대해 적분한 값은 1이다. 연속형 확률분포의 대표적인 예로는 정규분포(normal distribution), 지수분포(exponential distribution), 균등분포(uniform distribution) 등이 있다. [1.3 기술통계: 표와 그래프의 작성]에서는 실제 관측된 자료를 기반으로 경험적 빈도분포를 다루었다. 만약 경험적 빈도분포가 수학적으로 정의된 특정 이론적 확률분포와 유사하다면, 이론적 확률분포를 이용하여 확률을 계산하거나 모집단에 대한 추론을 할 수 있다.

2. 정규분포

정규분포(normal distribution)는 통계학에서 중요한 연속확률분포로 많은 통계 이론에 근간이 되는 확률분포이다. 정규분포의 모양은 평균 근처에 대다수의 값이 밀집되어 있고 평균에서 멀어질수록 값들의 밀집도가 감소하는 종 모양이다. 정규분포는 평균 μ와 표준편차 σ로 분포가 결정되며, 평균이 μ이고 표준편차가 σ인 표준정규분포를 $N(\mu, \sigma^2)$으로 표현한다. 정규분포의 확률밀도함수는 아래와 같다.

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ -\infty<x<\infty$$

표준편차가 동일한 σ이고, 평균이 각각 μ_1과 μ_2 ($\mu_1<\mu_2$)인 두 분포는 그림 1-1-3의 (a)와 같이 퍼짐정도는 동일하지만, 평균이 μ_2인 분포가 더 오른쪽에 분포한다. 그림 1-1-3의 (b)는 평균이 동일한 μ이지만, 표준편차가 σ_1, σ_2, σ_3 ($\sigma_1<\sigma_2<\sigma_3$)인 세 개의 정규분포를 나타낸다. 이 경우, 평균 위치는 같지만 퍼짐정도가 다르며, 표준편차가 커질수록 분포가 더 납작하다.

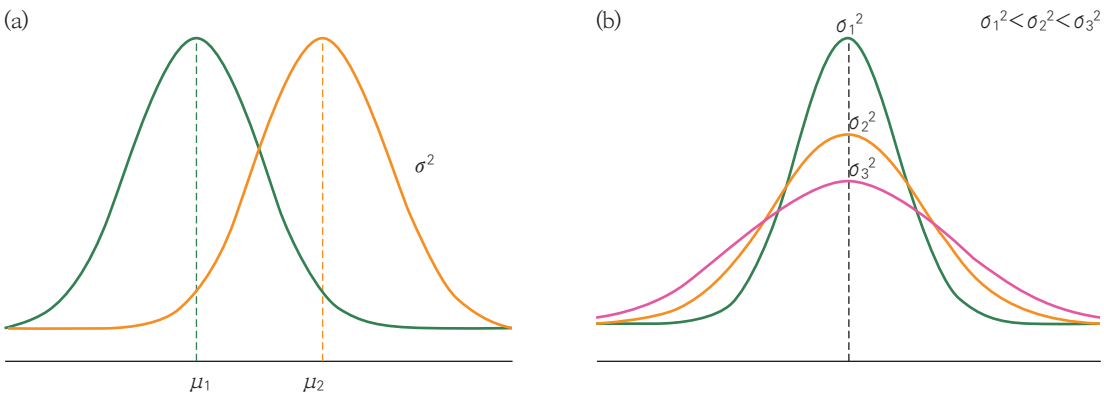


그림 1-1-3. 평균과 표준편차에 따른 정규분포의 형태

정규분포에서 확률을 구하는 방법은 확률밀도함수의 아래 면적을 계산하는 것과 같다. 이 면적은 확률밀도함수를 적분하여 계산할 수 있다. 예를 들어, 평균이 μ, 분산이 σ_2인 정규분포를 따르는 확률변수 X에 대해 $P(a<X<b)$는 확률밀도함수의 아래 면적(그림 1-1-4)과 같다.

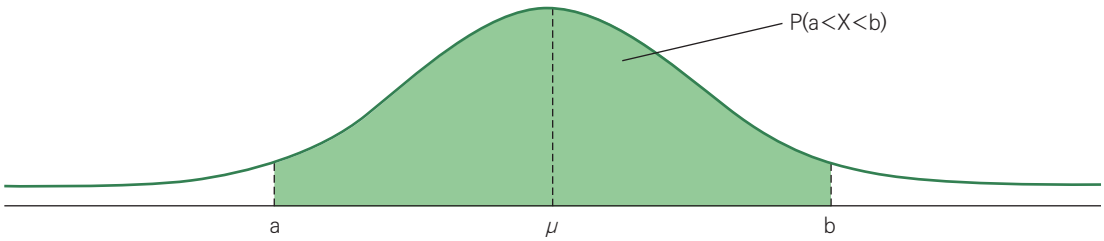


그림 1-1-4. 정규분포에서 구간 내 확률

평균이 μ이고 표준편차가 σ인 정규분포는 $z=\frac{x-\mu}{\sigma}$ 변환을 하면 평균이 0이고 표준편차가 1인 표준정규분포(standard normal distribution)가 된다. 이를 정규분포의 표준화라고 한다.

표준정규분포에서 Z값이 -1.96에서 1.96사이 면적은 95%이다. 즉, $P(-1.96\leq Z\leq 1.96)=0.95$로 확률변수 Z가 -1.96에서 1.96 사이에 존재할 확률이 95%임을 의미한다. 95% 확률을 가지는 Z값은 신뢰구간이나 가설검정에서 자주 사용되므로 기억해두자.

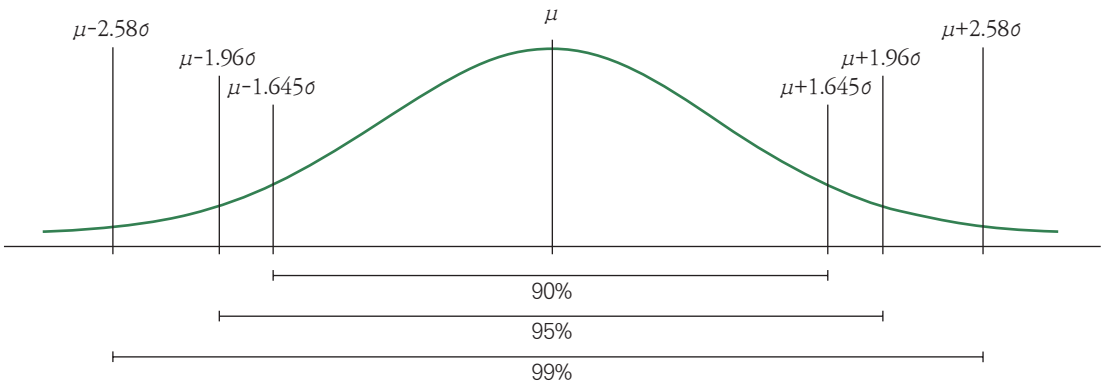


그림 1-1-5. 표준정규분포의 확률

3. 표본분포

표본분포(sampling distribution)란 모집단에서 동일한 크기의 표본을 여러 번 무작위로 추출했을 때, 각 표본에서 계산된 통계량들이 이루는 분포를 의미한다. 예를 들어, 표본크기가 n인 표본을 모집단에서 추출했을 때 다음은 자주 사용되는 통계량이다.

$$\text{표본평균: } \bar{x}=\frac{x_1+x_2+\cdots x_n}{n}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

$$\text{표본 분산: } S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

$$\text{표본 표준편차: } S=\sqrt{S^2}$$

표본크기가 n인 표본을 모집단에서 반복적으로 추출하고, 각 표본에서 특정한 통계량을 계산하면 표본오차로 인해 서로 다른 값들이 얻어진다. 이렇게 계산된 통계값들은 특정한 패턴을 따르는 분포를 형성하는데, 이를 표본분포라고 한다. 표본분포가 이론적 확률분포를 따른다면, 모집단 전체를 조사하는 대신 표본을 1회 추출하여 표본분포의 특성을 이용해 확률을 계산하거나 가설검정을 할 수 있다. 이 절에서는 표본평균의 표본분포와 표본분산의 표본분포, 표본비율의 표본분포에 대해 살펴본다.

1) 표본평균에 대한 표본분포

동일한 크기의 표본을 모집단에서 반복적으로 추출할 때, 각 표본에서 계산된 추정치는 항상 동일한 것은 아니다. 모집단에서 표본크기가 n인 표본 하나를 뽑아서 평균을 계산한 값을 $\bar{x}_1$이라고 하자. 그리고 동일한 모집단에서 표본크기가 n인 또 다른 표본 하나를 뽑아서 평균을 계산한 값을 $\bar{x}_2$라고 하자. 표본 추출 과정에서 발생하는 변동으로 인해 $\bar{x}_1$과 $\bar{x}_2$는 서로 다른 값을 가질 수 있다. 이와 같이, 동일한 크기의 표본을 모집단에서 반복적으로 추출하고 각 표본의 평균을 계산하면, 이러한 표본평균들의 값은 특정한 분포를 따르게 된다. 이 분포를 표본평균의 표본분포라고 한다.

표본분포는 반복적으로 표본을 추출하여 얻은 추정치들의 변동을 정량화함으로써, 추정치의 정밀도에 대한 정보를 제공하고 표본추출로 인한 오차를 평가할 수 있게 한다. 하지만 실제로 우리는 표본을 단 한 번만 추출하며, 이때 통계량의 이론적 분포에 대한 지식을 활용하면 모수에 대한 추론이 가능하다.

평균이 μ이고 표준편차가 σ인 모집단에서 표본크기가 n인 표본을 무작위로 반복추출하여 얻은 표본평균의 표본분포는 다음과 같은 특성을 갖는다.

$\bar{x}$의 표본분포:

- 표본평균의 평균은 모평균 μ와 같다.
- 표본평균의 표준편차는 $\sigma/\sqrt{n}$이다. 이때 $\sigma/\sqrt{n}$을 표준오차(standard error)라고 부른다.
- 모집단이 정규분포를 따르는 경우, 표본평균도 정규분포를 따른다.
- 모집단이 정규분포를 따르지 않는 경우, 중심극한정리(central limit theorem, CLT)에 의하여 표본크기 n이 충분히 크면 표본평균은 정규분포를 따른다.

위 내용을 종합하면 표본평균은 모평균 근처에 종 모양으로 분포하는데, 표본크기가 클수록 모평균 근처에 더 조밀하게 집중된다. 표본크기가 클수록 표준오차 $\sigma/\sqrt{n}$은 작아지는데 이는 추정치가 정밀하다는 의미이다. 반대로 표준오차가 크다는 것은 추정치가 정밀하지 않다는 의미이다. 이와 같이 표준오차는 표본평균의 정밀도에 대한 척도이다.

중심극한정리(central limit theorem, CLT)는 표본크기 n이 충분히 클 때 모집단의 분포와 상관없이, 표본평균의 분포는 정규분포에 근사한다는 것이다. 이는 모집단이 정규분포가 아닐지라도 표본크기가 커지면 표본평균의 분포가 정규분포에 가까워진다는 것을 의미한다.

중심극한정리

평균이 μ, 분산이 σ^2인 모집단에서 표본크기가 n인 표본을 추출했을 때 표본평균 $\bar{X}$의 분포는 n이 충분히 크면, 정규분포 $N\left(\mu, \frac{\sigma^2}{n}\right)$을 근사적으로 따른다.

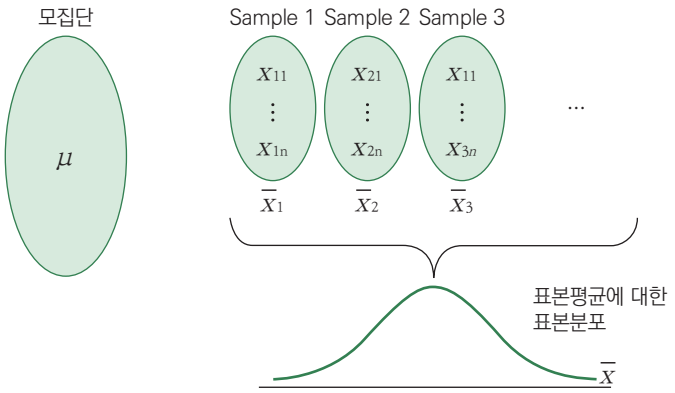


그림 1-1-6. 표본크기 n인 표본을 뽑았을 때 표본평균에 대한 표본분포

n이 충분히 크다는 것의 기준은 모집단의 분포가 정규분포에서 벗어난 정도에 따라 다르다. 모집단의 분포가 정규분포에 가까울수록 적은 표본으로도 표본평균의 분포는 정규분포를 따른다. 경험적으로 알려진 기준은 표본크기가 30이상일 때, 정규분포를 가정하더라도 무리가 없다고 알려져 있다.

2) 표본분산에 대한 표본분포

정규분포를 따르는 모집단에서 추출한 크기가 n인 표본을 $\{X_1, X_2, \cdots, X_n\}$이라고 하자. 표본분산 S^2의 표본분포는 카이제곱 분포를 따른다. 특히, 표본분산을 모분산으로 나누면, 자유도가 $n-1$인 카이제곱분포를 따른다. 카이제곱 분포는 왼쪽 꼬리가 긴 치우친 모양이며, 자유도가 커질수록 분포의 형태가 정규분포에 가까워지는 특성이 있다.

$$\frac{(n-1)S^2}{\sigma^2} \sim x^2(n-1)$$

3) 표본비율에 대한 표본분포

모비율(population proportion) p는 모집단에서 특정 특성을 가진 개체의 비율을 의미하는 모수

이다. 범주형 변수에서 모집단의 특성을 설명하기 위해 자주 사용되는 모수이다. 예를 들어, 우리나라의 백신 접종률은 전체 국민 중 백신을 맞은 사람의 비율로 정의할 수 있다. 표본비율 $\hat{p}=\frac{x}{n}$은 표본에서 특정 범주에 대한 빈도를 표본크기로 나눠서 계산한다. 표본비율은 다음과 같은 성질이 있다.

$$\text{표본비율 } \hat{p}\text{의 기댓값: } E(\hat{p})=p$$

$$\text{표본비율 } \hat{p}\text{의 분산: } Var(\hat{p})=\frac{p(1-p)}{n}$$

$$\text{표본비율 } \hat{p}\text{의 표준오차: } SE(\hat{p})=\sqrt{\frac{p(1-p)}{n}}$$

$np>5$이고 $n(1-p)>5$일 때, $\hat{p}$의 표본분포가 정규분포에 근사한다.

4. 점추정과 구간추정

추정(estimation)이란 표본의 통계량을 활용하여 모집단의 모수를 추정하는 과정을 의미한다. 모집단의 평균이나 비율과 같은 모수는 중요한 정보를 담고 있다. 예를 들어, "대한민국 고령자의 독감 접종률"은 보건 정책 수립과 의사결정에 핵심적인 근거로 활용될 수 있다. 하지만 모집단은 대개 무한모집단이거나, 유한모집단이라 하더라도 규모가 매우 커서 전수 조사를 수행하는 것은 현실적으로 불가능한 경우가 많다. 따라서 표본을 추출하여 표본으로부터 통계량을 계산하고 이를 이용하여 모수를 추정하는 것이 바람직하다. 이때 모수를 추정하기 위해 사용되는 통계량을 추정량(estimator)이라고 한다.

추정은 점추정(point estimation)과 구간추정(interval estimation) 두 가지 방법이 있다. 점추정은 모집단의 모수를 단일 값으로 추정하는 것을 의미하며, 간단하고 직관적인 방법이다. 반면, 값을 특정 범위 형태로 나타내는 방법은 구간추정이라고 한다. 앞서 다룬 표준오차는 점추정치의 정밀도에 대한 정보를 제공하지만, 단순히 그 자체로 사용되기보다는 모수에 대한 구간추정치를 계산하는데 활용된다. 구간추정에서 주로 사용하는 방법은 신뢰구간(confidence interval)이다. 신뢰구간은 모수의 값이 포함되리라고 기대되는 범위를 의미한다. 점추정은 계산이 간단하지만, 내가 추출한 하나의 표본을 통해 얻어진 점추정치가 실제 모수, 예를 들어 모평균과 비슷할 것인지 확신하기 어렵다. 이 불확실성을 보완하기 위해 신뢰구간을 활용하여 점추정치의 신뢰도를 범위로 표현한다. 신뢰구간은 점추정치에 대한 불확실성을 수치화하여 보다 구체적인 추론을 가능하게 한다.

신뢰구간은 점추정량를 중심으로 좌우로 표준오차의 신뢰계수 배만큼 떨어진 값들을 계산하여 구간

의 형태로 나타낸 것이다. 이때 신뢰계수는 점추정량의 분포와 구하려는 신뢰수준에 따라 결정된다. 신뢰수준은 보통 95%를 많이 사용한다.

점추정량±(신뢰계수)×(표준오차)

1) 모평균 μ에 대한 신뢰구간

모평균 μ에 대한 신뢰구간을 구하기 위해 표본평균 $\overline{X}$의 표본분포를 이용하여 신뢰구간을 구해보자. 앞서 설명한 것처럼, 평균이 μ이고 표준편차가 σ인 모집단에서 표본을 추출하면, 표본평균의 표본분포는 평균이 μ이고 표준편차가 $\sigma/\sqrt{n}$인 정규분포를 따른다. 모평균에 대한 신뢰구간을 구할 때는 이러한 표본평균의 표본분포의 특성을 활용한다. 정규분포의 특성에 따르면, 표본평균들 중 95%는 모평균 μ를 중심으로 좌우 1.96배 표본평균의 표준편차(표준오차, standard error, SE) 범위 내에 존재하게 된다. 여기서 1.96는 $z_{1-\alpha/2}$(α=0.05)에 해당하는 값으로 95% 신뢰구간에 대한 신뢰계수이다.

$$\mu-1.96\times\frac{\sigma}{\sqrt{n}}<\bar{x}<\mu+1.96\times\frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow \bar{x}-1.96\times\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+1.96\times\frac{\sigma}{\sqrt{n}}$$

이때 모평균에 대한 신뢰구간은 다음과 같이 계산된다:

$$\left(\bar{x}-1.96\times\frac{\sigma}{\sqrt{n}},\ \bar{x}+1.96\times\frac{\sigma}{\sqrt{n}}\right)$$

신뢰구간은 많은 사람들이 모평균이 포함될 확률을 의미하는 것으로 오해하지만, 모평균은 확률변수가 아니라 모수로서 정해진 값이므로 특정 구간에 포함될 확률을 갖지 않는다. 그렇다면 신뢰구간은 어떤 의미를 가지는가? 그림 1-1-7과 같이, 동일한 표본크기의 표본을 100번 추출하여 각 표본에서 신뢰구간 $\bar{x}\pm z_{a/2}\frac{\sigma}{\sqrt{n}}$을 계산하면 100개의 신뢰구간을 얻을 수 있다. 이때, 이 신뢰구간 중 100(1-α)%는 모평균 μ를 포함하게 된다. 예를 들어, 95% 신뢰구간이라면, 100개의 신뢰구간 중 약 95개는 모평균을 포함하며, 나머지 5개는 포함하지 않을 가능성이 있다. 즉, 신뢰구간은 표본을 반복적으로 추출하여 각 표본에서 신뢰구간을 계산하면, 이들 중 100(1-α)%는 모수를 포함한다는 개념이다.

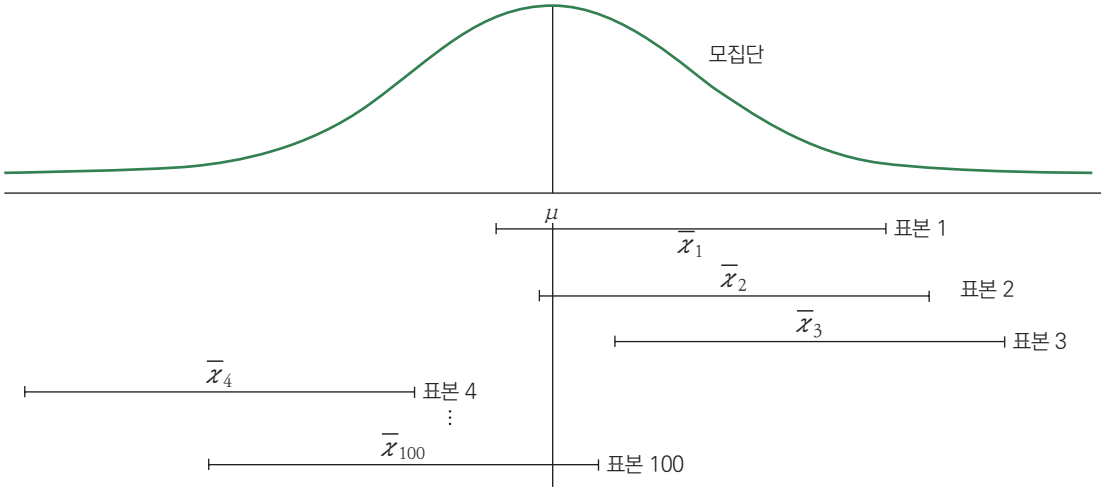


그림 1-1-7. 표본평균에 대한 표본분포와 μ의 95% 신뢰구간

(1) 모분산이 알려지지 않은 경우 모평균 μ의 신뢰구간(대표본)

모분산이 알려지지 않은 경우에는, 표본크기 n이 크면 표준오차 $\sigma/\sqrt{n}$ 대신 표준오차의 추정량으로 대체할 수 있다. 표준오차의 추정량은 모표준편차 σ를 대신하여 표본 표준편차를 사용한 $s/\sqrt{n}$이다. 모분산이 알려지지 않은 경우 n이 크다면 모평균 μ에 대한 $100(1-\alpha)\%$ 신뢰구간은 다음과 같다.

$$\left(\bar{x}-1.96\times\frac{s}{\sqrt{n}},\ \bar{x}+1.96\times\frac{s}{\sqrt{n}}\right)$$

(2) 모분산이 알려지지 않은 경우 모평균 μ의 신뢰구간(소표본)

이때, 표본평균의 표본분포는 정규분포가 아닌 t-분포를 따른다. t-분포는 자유도(degree of freedom, df)에 따라 형태가 달라지며, 자유도가 커질수록 정규분포에 근사한다. 또한, t-분포는 평균 0을 중심으로 좌우 대칭이며, 정규분포보다 꼬리가 두꺼운 형태를 가진다.

표본평균의 표본분포가 t-분포를 따른다는 점을 이용하여 모평균 μ의 신뢰구간을 구할 수 있으며, 신뢰계수로 $t_{1-\alpha/2}(n-1)$을 사용한다. 여기서 $t_{1-\alpha/2}(n-1)$는 t-분포에서 $t_{1-\alpha/2}(n-1)$보다 작을 확률이 $1-\alpha/2$일 t값을 의미한다. 즉, $P(T\leq t_{1-\alpha/2}(n-1))=1-\alpha/2$이다. 결과적으로 모분산이 알려지지 않은 경우 표본크기 n이 크지 않다면, 모평균 μ에 대한 $100(1-\alpha)\%$ 신뢰구간은 다음과 같다.

$$\left(\bar{x}-t_{1-\alpha/2}(n-1)\frac{s}{\sqrt{n}},\quad \bar{x}+t_{1-\alpha/2}(n-1)\frac{s}{\sqrt{n}}\right)$$



그림 1-1-8. 정규분포와 t-분포의 비교

2) 모비율 p에 대한 신뢰구간

표본비율의 표본분포는 $np>5$이고 $n(1-p)>5$일 때, 정규분포에 근사한다고 하였다. 이를 바탕으로 신뢰구간은 표준정규분포를 이용하여 구할 수 있다. 표본비율의 표준오차는 $\sqrt{\frac{p(1-p)}{n}}$이며, 표준오차의 추정량은 $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$이다. 따라서 모비율 p에 대한 $100(1-\alpha)\%$ 신뢰구간은 다음과 같다. 여기서 $z_{1-\alpha/2}$는 $p(Z\leq z_{1-\alpha/2})=1-\alpha/2$을 만족하는 표준정규분포의 Z-값이다.

$$\left(\hat{p}-z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\quad \hat{p}+z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

5 통계적 가설검정

1. 가설검정의 개념

이 장에서는 통계적 추론의 핵심 개념인 가설검정에 대해 학습하며, 기초 개념부터 다양한 검정 방법까지 익히고, 이를 실제 데이터 분석에 적용하는 방법을 배운다. 통계적 가설검정(statistical hypothesis testing)은 표본 자료를 통해 모집단의 특성에 대한 가설(hypothesis)의 타당성을 평가하는 과정이다.

표본 자료를 수집하는 이유 중 하나는 모집단에 대해 설정한 특정 가설에 반하는 증거를 얼마나 담고 있는지 평가하기 위함이다. 표본 자료에 대한 기술통계(descriptive statistics) 분석만 수행해도 해당 자료가 가설을 지지하는 경향이 있는지, 아니면 이에 반하는 경향이 있는지 어느 정도 파악이 가능하다. 그러나 기술적 분석만으로는 표본 자료의 패턴이 실제로 변수 간 상관성(correlation)을 반영하는 것인지, 아니면 단순히 우연한 변동(variation)에 의한 것인지 구분하기 어렵다.

가설검정은 표본 자료와 귀무가설을 바탕으로 계산된 검정통계량이 귀무가설 하에서 확률적으로 일어나기 매우 어려운 것이라면 귀무가설을 기각한다. 반면, 검정통계량이 귀무가설 하에서 극단적인 결과가 아니라면 기각하지 않는 방식으로 수행한다. 즉, 가설검정은 표본 자료를 기반으로 모집단에 대한 결론을 도출하는 방법으로, 단순한 기술적 분석을 넘어 확률적인 근거를 바탕으로 가설의 타당성을 검토하는 과정이다.

1) 귀무가설과 대립가설

가설검정에서 가설에는 귀무가설(null hypothesis)과 대립가설(alternative hypothesis)이 있다. 귀무가설은 일반적으로 받아들여지는 사실로 H_0로 표시한다. 반면, 대립가설이란 연구자가 연구로 밝혀내고자 하는 가설로 H_1으로 표시한다. 예를 들어, 특정 지역에서 실시한 대사증후군 관리 사업이 혈압을 낮추는데 효과가 있는지 확인하고자 한다고 가정하자. 여기서 대립가설은 "대사증후군 관리 사업이 혈압을 낮추는데 효과가 있다"이며, 귀무가설은 대립가설에 반하는 주장인 "대사증후군 관리 사업이 혈압을 낮추는데 효과가 없다"가 된다.

보통 귀무가설은 "차이가 없다", "효과가 없다"와 같은 형태로 설정된다. 대립가설은 "차이가 있다", "효과가 같지 않다", "효과가 더 좋다"와 같은 형태로 설정되는데, 이때 "같지 않다"는 양측검정(two-sided test)라고 하고, "더 좋다"와 같이 어느 한 방향으로 대립가설이 설정되는 것을 단측검정(one-sided test)라고 한다. 다음은 양측검정과 단측검정의 예시이다.

양측검정

- H_0: 남자와 여자의 흡연율은 같다. ($p_1=p_2$)
- H_1: 남자와 여자의 흡연율은 같지 않다. ($p_1 \neq p_2$)

단측검정

- H_0: 남자와 여자의 흡연율은 같다. ($p_1=p_2$)
- H_1: 남자의 흡연율이 여자의 흡연율보다 크다. ($p_1>p_2$)

2) 의사결정의 오류: 제1종 오류, 제2종 오류, 검정력

가설검정 시 발생하는 의사결정 오류는 제1종 오류(type I error)와 제2종 오류(type II error) 두 가지가 있다. 제1종 오류는 귀무가설이 사실인데도 불구하고 이를 기각하는 오류이다. 이때, 제1종 오류가 일어날 확률의 최대허용치를 유의수준(significance level)이라고 하며, α로 표기한다. 반면 제2종 오류는 귀무가설이 거짓임에도 불구하고 이를 기각하지 않는 오류를 의미하며, 제2종 오류에 대한 확률은 β로 표기한다. 거짓인 귀무가설을 기각하게 될 확률은 $1-\beta$이고 이를 검정력이라고 한다. 검정력의 크기를 결정하는 요인은 표본수, 관찰값들의 변동량, 효과크기, 유의수준이 있다. 일반적으로 표본수가 클수록 검정력이 증가하며, 관찰값들의 변동이 작을수록 검정력이 높아진다. 또한, 처리의 효과크기가 클수록 검정력이 증가하며, 유의수준이 높을수록 검정력도 커지는 관계가 있다.

예를 들어, 특정 처리(treatment)의 효과에 대해 가설검정을 한다고 하자. 이때, 귀무가설은 "처리의 효과가 없다"이다. 만약 실제로 처리의 효과가 없는데도 불구하고 효과가 있다고 판단하면 이는 제1종 오류에 해당한다. 반대로, 실제로 처리의 효과가 존재하지만 효과가 없다고 결론을 내리면 이는 제2종 오류에 해당한다.

가설검정에서 발생하는 의사결정의 오류

검정 결과	모집단에서	
	귀무가설(H_0) 참	귀무가설(H_0) 거짓
귀무가설(H_0) 기각	제 1종 오류(type Ⅰ error)	옳은 결정
귀무가설(H_0) 기각하지 않음	옳은 결정	제 2종 오류(type Ⅱ error)

두 가지 오류 모두 작을수록 좋지만, 제1종 오류를 줄이면 제2종 오류가 커지고, 반대로 제2종 오류를 줄이면 제1종 오류가 커지는 특성이 있다. 따라서 일반적인 가설검정에서는 제1종 오류의 최대허용치인 유의수준을 고정하고, 주어진 유의수준에서 검정력을 최대화하는 검정법을 선택한다. 보통 유의수준은 5%로 설정하지만, 상황에 따라서 1% 또는 10%로 정하기도 한다. 만약 귀무가설을 잘못 기각하는 경우(제1종 오류)가 귀무가설을 잘못 채택하게 되는 경우(제2종 오류)보다 더 심각한 결과를 초래하

는 상황이라면, 유의수준을 낮게 설정한다. 반대로, 제2종 오류를 줄이는 것이 더 중요한 경우에는 유의수준을 10%로 설정하기도 한다.

3) p-value

p-value (유의확률)는 귀무가설이 참이라는 가정하에, 표본 자료로부터 얻어진 통계량이 더 극단적인 값을 가지게 될 확률을 말한다. 즉, "효과가 없다"는 귀무가설이 참임에도 불구하고 우연히 통계량 이상의 값이 관측될 확률을 계산하는 것이다. 예를 들어, p-value가 0.3으로 도출되었다고 하자. 이는 귀무가설이 참이라는 가정하에 더 극단적인 결과가 나올 확률이 약 3%라는 의미이다. 다시 말해, 100번의 실험에서 3번 정도는 귀무가설이 참일 수 있다고 해석할 수 있다. 만약 p-value가 매우 작다면, 현재의 자료가 귀무가설 하에서 관찰될 가능성이 낮다는 것을 의미하며, 이는 귀무가설을 기각할 강한 증거가 된다. 이때 통계량 이상이 관측될 확률은 검정통계량(test statistics)의 확률분포를 이용하여 계산한다. p-value는 이 확률분포에서 검정통계량에 해당하는 값보다 극단적인 영역의 면적을 의미한다. 양측검정의 경우, 검정통계량 값의 양쪽 극단에 해당하는 면적이며, 단측검정의 경우 한쪽 극단의 면적이다.

그렇다면 p-value가 얼마나 작아야 귀무가설을 기각할까? 일반적으로 p-value가 미리 설정한 유의수준(α)보다 작으면 귀무가설을 기각한다. 만약 유의수준을 5%로 설정하였다면, p-value가 0.05보다 작을 때 귀무가설을 기각할 충분한 근거가 있다고 본다. 이를 "유의수준 5%에서 유의하다."라고 표현한다.

p-value의 해석에서 주의점은 그것이 귀무가설의 참/거짓을 직접적으로 증명하는 것이 아니라, 단순히 귀무가설 하에서 표본 자료의 통계량보다 극단적인 결과가 나올 확률을 제공하는 것이라는 점이다. 따라서 p-value가 작다고 해서 반드시 대립가설이 참이라고 단정할 수는 없다.

4) 가설검정의 절차

가설검정을 일반적으로 다음과 같은 단계로 수행한다.

가설검정의 절차

1. 귀무가설과 대립가설을 설정한다.
2. 표본으로부터 자료를 수집한다.
3. 적절한 유의수준 α를 설정한다.
4. 귀무가설에 해당하는 검정통계량 값과 p-value를 계산한다.

5. p-value와 유의수준을 비교하여 해석하고 결론을 내린다.

5단계에서 결론을 내릴 때 검정통계량과 임곗값(critical value)를 비교하는 방법과 p-value와 유의수준을 비교하는 방법 두 가지가 있으나, 이 두 방법은 본질적으로 동일한 결과를 제공한다.

2. 연속형 자료에 대한 두 집단의 비교

이 절에서는 두 집단 간 연속형 자료를 비교하는 방법에 대해 다룬다. 이때, 독립표본과 짝지은 표본의 가설검정 방법이 서로 다르므로, 먼저 독립적인 두 집단의 비교 방법을 살펴보고 이후 짝지은 표본에서의 비교 방법을 알아볼 것이다. 검정방법은 연구의 설계, 변수의 종류, 연구에서 얻어진 자료가 따르는 분포의 형태 등에 따라 결정된다. 두 집단 간 비교 검정 방법에 대한 대략적인 분석 흐름은 그림 1-1-9와 같다. 이 절에서는 정규분포를 가정한 모수적(parametric) 방법을 중심으로 설명한다. 두 모평균 차이에 대한 검정 방법은 다음 세 가지 경우에 따라 달라진다.

- 모집단이 정규분포를 따르며, 모분산이 알려져 있는 경우 → z-검정
- 모집단이 정규분포를 따르며, 모분산이 알려지지 않은 경우 → 모분산의 동질성 검정 후 t-검정
- 모집단이 정규분포를 따르지 않는 경우 → 비모수 검정

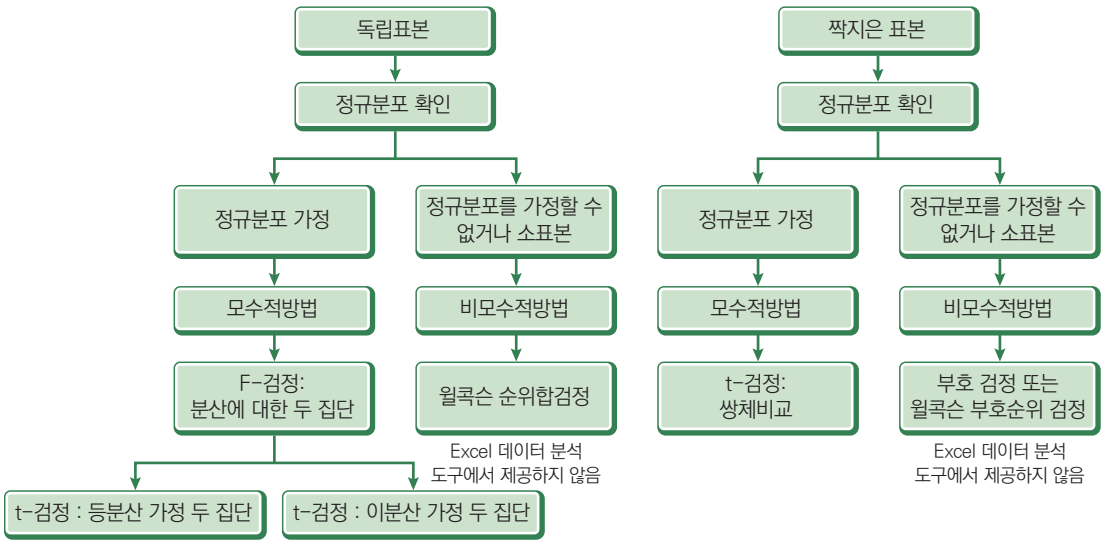


그림 1-1-9. 두 집단 간 연속형 자료의 비교

따라서 모집단에 대한 분포를 먼저 확인한 후, 정규분포를 가정할 수 있는 경우에는 모수적 방법을 사용하고, 정규분포를 가정하기 어려운 경우에는 비모수 검정을 대안으로 수행한다.

1) 정규분포 확인

(1) 정규성 검정

주어진 표본 자료가 정규분포를 따르는 모집단에서 랜덤하게 추출된 것인지 확인하기 위해 수행하는 가설검정을 정규성 검정(test for normality)라고 한다. 대표적인 정규성 검정 방법으로 콜모고로프-스미노프 검정(Kolmogorov-Smirnov test)과 샤피로-윌크 검정(Shapiro-Wilk test)이 있다. 콜모고로프-스미노프 검정은 표본크기가 충분히 큰 경우에 적합한 검정 방법으로 알려져 있다. 샤피로-윌크 검정은 표본크기가 비교적 작을 때 추천되는 방법이다. 다만, 표본크기에 따라 검정 방법의 적합성을 나누는 것은 편의성을 위한 절차일 뿐, 절대적인 것은 아니다. 따라서 그래프를 활용하여 분포를 시각적으로 확인하고, 정규성 검정 결과와 함께 종합적으로 판단하는 것이 중요하다.

정규성 검정의 가설은 다음과 같다.

- H_0: 자료는 정규분포를 따르는 모집단에서 임의로 추출된 자료이다.
- H_1: 자료는 정규분포를 따르는 모집단에서 임의로 추출된 자료가 아니다.

(2) 그래프를 이용한 정규분포 확인

정규성을 그래프로 확인하는 방법은 데이터가 정규분포를 따르는지 시각적으로 평가하는 데 사용된다. 주요 그래프와 방법은 다음과 같다.

① 히스토그램

히스토그램(histogram)은 데이터의 분포를 시각적으로 보여준다. 데이터가 정규분포를 따른다면 히스토그램의 모양은 중심이 높고 양쪽으로 대칭적인 종 모양(bell shape)을 이룬다.

② Q-Q 플롯

Q-Q 플롯(Quantile-Quantile Plot)은 데이터의 분포와 정규분포의 이론적 분포를 비교하는 그래프이다. x축은 정규분포의 이론적 분위 값을 나타내고, y축은 실제 데이터의 분위 값을 나타낸다. 데이터가 정규분포를 따른다면 Q-Q 플롯의 점들이 대각선 직선에 가깝게 배열된다. 점들이 직선에서 크게 벗어나면 정규성을 만족하지 않을 가능성이 높다.

③ P-P 플롯(Probability-Probability Plot)

P-P 플롯은 Q-Q 플롯과 유사하게 데이터를 정규분포의 누적확률과 비교한다. 데이터의 누적확률 값이 정규분포의 누적확률 값과 일치하면 직선 형태를 이룬다.

2) 두 독립 집단의 평균비교

역학 연구에서는 대조군과 실험군을 비교하는 경우가 많다. 이 절에서는 비교하고자 하는 결과변수가 연속형일 때 두 집단의 평균을 비교하는 방법에 대해 알아보자. 두 집단의 평균을 비교할 때 충족해야 하는 전제는 두 집단이 서로 독립적인(서로 관련이 없는) 모집단에서 추출된 표본이어야 한다는 것이다. 또한, 관심 변수는 수치형 혹은 순서형 변수인 경우가 일반적이다. 두 모평균 차이 δ에 대한 검정에서 다음과 같은 가설들을 설정할 수 있다.

- 양측검정: H_0: $\mu_1-\mu_2=\delta$ vs H_1: $\mu_1-\mu_2\neq\delta$
- 단측검정: 좌측검정 H_0: $\mu_1-\mu_2=\delta$ vs H_1: $\mu_1-\mu_2<\delta$
- 단측검정: 우측검정 H_0: $\mu_1-\mu_2=\delta$ vs H_1: $\mu_1-\mu_2>\delta$

(1) 두 독립 집단의 평균비교: 모분산이 알려진 정규모집단의 경우

모집단이 정규분포를 따르며, 모분산이 알려져 있는 경우에 두 집단 간의 모평균 차이에 관한 가설검정에 대해 알아보자.

① 가정

- $x_1, x_2, \cdots, x_{n_1}$은 평균이 μ_1이고, 분산이 σ_1^2인 정규분포를 따르는 모집단에서 추출된 표본이다.
- $y_1, y_2, \cdots, y_{n_2}$는 평균이 μ_2이고, 분산이 σ_2^2인 정규분포를 따르는 모집단에서 추출된 표본이다.
- 두 표본은 서로 독립이다.

② 검정통계량

H_0: $\mu_1-\mu_2=\delta$를 검정하기 위한 검정통계량 z는 아래와 같다. 각 표본이 서로 다른 정규 모집단에서 추출되었다고 가정하였으므로, 표본평균 $\bar{x}$, $\bar{y}$의 분포와 표본평균의 차이($\bar{x}-\bar{y}$)의 분포는 다음과 같다.

$$\bar{x}\sim N(\mu_1,\ \sigma_1^2/n_1),\ \bar{y}\sim N(\mu_2,\ \sigma_2^2/n_2)$$
$$\rightarrow \mathrm{x\text{-}y}\sim N(\mu_1-\mu_2,\ \sigma_1^2/\mathrm{n}_1+\sigma_2^2/n_2)$$

검정통계량 Z는 다음과 같이 정의되며, 귀무가설 하에서 표준정규분포를 따른다.

$$Z=\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/\mathrm{n}_2}}=\frac{(\bar{x}-\bar{y})-\delta}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/\mathrm{n}_2}}\sim N(0,1)$$

이를 이용하여 유의수준을 α로 설정하였을 때 가설에 따른 결과는 다음과 같다.

- 양측검정: H_0: $\mu_1-\mu_2=\delta$ vs H_1: $\mu_1-\mu_2\neq\delta$ — $|Z|>z_{1-\alpha/2}$이면 귀무가설 기각
- 단측검정: 좌측검정 H_0: $\mu_1-\mu_2=\delta$ vs H_1: $\mu_1-\mu_2<\delta$ — $Z<z_{\alpha}$이면 귀무가설 기각
- 단측검정: 우측검정 H_0: $\mu_1-\mu_2=\delta$ vs H_1: $\mu_1-\mu_2>\delta$ — $Z>z_{1-\alpha}$이면 귀무가설 기각

이때 모평균차이의 100(1-α)% 신뢰 구간은 다음과 같다.

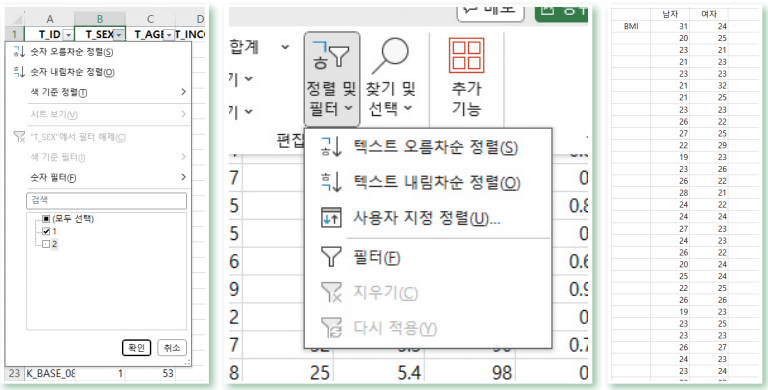
③ 모평균차이 $\mu_1-\mu_2$의 100 (1−α)% 신뢰구간

$$\left(\bar{x}-\bar{y}-Z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\ \bar{x}-\bar{y}-Z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

[실습] 모분산이 알려진 경우, 성별에 따른 BMI의 모평균 차이 비교(ex_survey_data.xlsx)

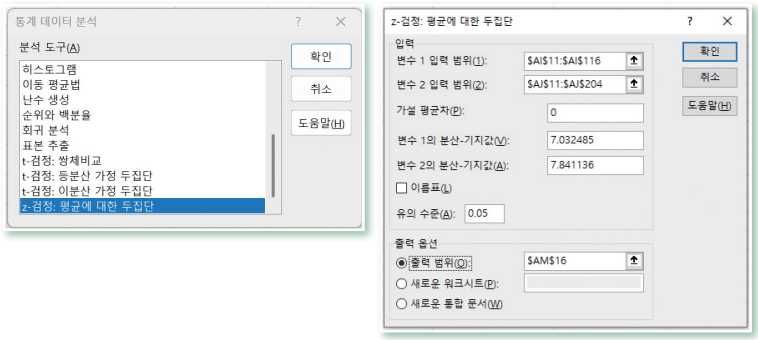
모분산이 알려진 정규모집단을 가정하고, 성별(남자, 여자)에 따라 BMI(체질량지수)의 모평균이 통계적으로 유의미한 차이가 있는지 확인하고자 한다. Excel의 데이터분석 도구에서 "Z-검정: 평균에 대한 두집단" 기능을 이용하여 검정을 수행해보자.

가. T_SEX 변수를 선택한 후, [홈] → [정렬 및 필터]를 통해 필터 기능을 사용한다. 성별이 1인 남성을 필터링 한 후, BMI의 수치를 복사하여, 새로운 시트에 데이터셋을 생성한다. 다음으로, 성별이 2인 여성을 필터링하여 같은 방법을 수행한다.



나. Z-검정은 모집단의 분산이 알려진 경우에 사용하는 통계 분석 방법이므로, 두 집단의 알려진 모분산값을 쓴다. 여기서는 모분산 대신 표본분산 7.032와 7.841 값을 넣었다.

다. [데이터] 탭 [분석] 그룹에서 데이터분석을 누른 후, z-검정: 평균에 대한 두 집단을 선택한다



- 변수 1 입력 범위에는 남성의 BMI, 변수 2 입력 범위에는 여성의 BMI를 지정한다.
- 귀무가설이 “두 집단의 평균에 유의미한 차이가 없다”이므로 가설의 평균차 δ는 “0”으로 쓴다.
- 분산-기지값은 모분산을 의미하므로, 이전에 계산한 모분산을 작성한다.
- 유의수준은 5%으로 작성하였고, 출력 범위를 빈 셀로 지정하여 결과를 확인한다.

라. [$P(Z \le z)$ 단측 검정]과 [$P(Z \le z)$ 양측 검정] 모두 유의수준 0.05 보다 작으므로, 귀무가설을 기각한다. 즉, 두 집단의 평균에 유의미한 차이가 있다.

z-검정: 평균에 대한 두 집단

	남성의 BMI	여성의 BMI
평균	24.33018868	23.58247423
기지의 분산	7.032485	7.841136
관측수	106	194
가설 평균차	0	
z 통계량	2.288371659	
P(Z<=z) 단측 검정	0.011057943	
z 기각치 단측 검정	1.644853627	
P(Z<=z) 양측 검정	0.022115887	
z 기각치 양측 검정	1.959963985	

(2) 모분산이 알려지지 않은 경우 두 집단의 분산 동질성 검정

현실적으로 모분산을 정확히 아는 경우는 드물기 때문에, 일반적으로 모분산이 알려지지 않은 상황에서 두 모집단의 평균 차이에 대한 가설검정 방법이 많이 활용된다. 이 경우, 모분산이 동일한 경우와 동일하지 않은 경우에 따라 검정통계량이 달라지므로, 이를 구분하여 적절한 검정 방법을 선택해야 한다. 이를 위해 먼저 모분산이 알려지지 않은 경우에는 두 집단의 모분산에 대한 비교 검정을 수행한 후 분산의 동질성 검정 결과에 따라 적합한 t-검정 통계량을 선택하여 두 집단의 평균 비교 검정을 한다.

① 가설

H_0: 두 집단의 모분산은 같다. ($\sigma_1^2/\sigma_2^2=1$)

H_1: 두 집단의 모분산은 다르다. ($\sigma_1^2/\sigma_2^2\neq1$)

② 가정

- $x_1, x_2, \cdots, x_{n_1}$은 평균이 μ_1이고, 분산이 σ_1^2인 정규분포를 따르는 모집단에서 추출된 표본이다.
- $y_1, y_2, \cdots, y_{n_2}$는 평균이 μ_2이고, 분산이 σ_2^2인 정규분포를 따르는 모집단에서 추출된 표본이다.
- 두 표본은 서로 독립이다.

③ 검정통계량

검정통계량은 귀무가설이 사실일 때 자유도가 (n_1-1, n_2-1)인 F분포를 따른다. 여기서 n_1과 n_2는 각각 두 표본의 표본크기이며, s_1^2과 s_2^2은 각각 두 표본의 분산이다.

$$F=\frac{s_1^2/s_2^2}{\sigma_1^2/\sigma_2^2}=\frac{s_1^2}{s_2^2}\sim F(n_1-1, n_2-1)$$

④ 가설검정 결과

$F<F_{\alpha/2}(n_1-1, n_2-1)$ 또는 $F>F_{1-\alpha/2}(n_1-1, n_2-1)$이면 귀무가설을 기각하며, 반대의 경우 기각할 수 없다. p-value에 기반한 결론은 다음과 같으며, 이는 임곗값과 검정통계량을 비교하는 방법과 동일한 결과를 제공한다.

- p-value<α이면 귀무가설을 기각하며, 두 집단 간 모분산이 같지 않다.
- p-value>α이면 귀무가설을 기각할 수 없으며, 두 집단 간 모분산은 같다.

분산의 동질성 검정 결과에 따라 모분산이 같다고 가정한 경우와 같지 않다고 가정한 경우에 t-검정에서 검정통계량이 달라진다. 이는 다음 절에서 설명한다. 한편, 분산의 동질성 검정 역시 모집단이 정규분포를 따른다는 가정을 하고 있는데, 모집단이 정규분포를 따른다는 증거가 부족할 경우 레빈 검정(Levene's test)을 통해 등분산성에 관한 가설검정을 수행할 수 있다.

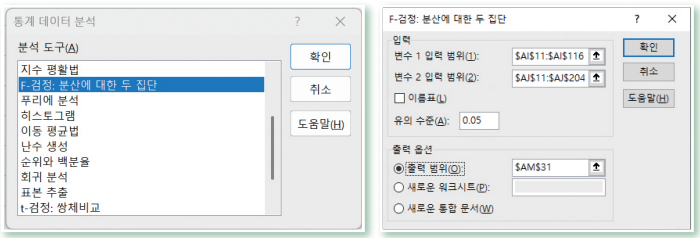
[실습] 남자와 여자의 BMI에 모분산 비교 검정

두 집단의 모분산을 비교하는 방법이다. z-검정의 예제와 같은 성별(남자, 여자)에 따라 BMI (체질량지수)의 평균이 통계적으로 유의미한 차이가 있는지 확인하고자 한다. 그 전에 분산의 동질성 검정을 수행해보자.

가. T_SEX 변수를 선택한 후, [홈] → [정렬 및 필터]를 통해 필터 기능을 사용한다. 성별이 1인 남성을 필터링 한 후, BMI의 수치를 복사하여, 새로운 시트에 데이터셋을 생성한다. 다음으로, 성별이 2인 여성을 필터링하여 같은 방법을 수행한다.

나. [데이터] 탭 [분석] 그룹에서 데이터분석을 누른 후, F-검정: 분산에 대한 두집단을 선택한다.

- 변수 1 입력 범위에는 남성의 BMI, 변수 2 입력 범위에는 여성의 BMI를 지정한다.
- 귀무가설은 "두 집단의 분산에 유의미한 차이가 없다."이다.
- 유의수준은 5%으로 작성하였고, 출력 범위를 빈 셀로 지정하여 결과를 확인한다.



다. [$P(F \le f)$ 단측 검정]이 유의수준 0.05 보다 큰 값이므로, 귀무가설을 기각하지 않는다. 즉, 두 집단의 분산에 유의미한 차이가 없다.

또한, [F-비] 0.90 이라는 값이 두 그룹의 분산 비율이 1에 가까운 값을 나타내므로, 두 그룹의 분산이 비슷하다는 것을 나타낸다.

F-검정: 분산에 대한 두 집단		
	남성의 BMI	여성의 BMI
평균	24.33018868	23.58247423
분산	7.099460916	7.881763795
관측수	106	194
자유도	105	193
F 비	0.900745201	
P(F<=f) 단측 검정	0.278151598	
F 기각치: 단측 검정	0.748267716	

(3) 두 독립 집단의 평균비교: 등분산 가정에서 독립 이표본 t-검정

두 독립된 집단의 모평균이 유의한 차이를 보이는지 검정하기 위해 표본수가 적으면서 두 모집단이 각각 정규분포를 따른다고 가정할 수 있을 때 독립 이표본 t-검정(independent two-sample t-test)이 가능하다. 독립 이표본 t-검정에서는 두 집단 간 모분산이 동일한 경우(등분산)와 동일하지 않은 경우(이분산)에 따라 검정 방법이 달라진다. 이 절에서는 두 집단 간 모분산이 같다는 가정(등분산 가정, homogeneity of variance) 하에서 독립 이표본 t-검정의 개념과 방법을 살펴본다. 두 모집단의 분산이 동일하다고 가정할 수 있다면, 두 집단의 표본분산은 두 표본의 표본분산을 병합한 합동분산(pooled variance)으로 추정할 수 있다. 합동분산은 다음과 같다.

$$s^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{(n_1+n_2-2)}$$

여기서 s_1^2과 s_2^2은 각각 두 집단의 표본분산이며, n_1과 n_2는 각각 두 집단의 표본크기이다.

① 검정통계량

H_0: $\mu_1-\mu_2=\delta$에 대한 검정통계량은 다음과 같으며, 귀무가설하에서 n_1+n_2-2인 t-분포를 따른다.

$$t=\frac{(\bar{x}-\bar{y})-\delta}{s\sqrt{(1/n_1+1/n_2)}}\sim t(n_1+n_2-2),\ \text{where}\ s=\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{(n_1+n_2-2)}}$$

② 가설검정 결과

유의수준 α로 설정했을 때 가설에 따른 검정 결과는 다음과 같다.

- 양측검정: H_0: $\mu_1-\mu_2=\delta$ vs H_1: $\mu_1-\mu_2\neq\delta$ $t>t_{1-\alpha/2}(n_1+n_2-2)$이면 귀무가설 기각

- 단측검정: 좌측검정 H_0: $\mu_1-\mu_2=\delta$ vs H_1: $\mu_1-\mu_2<\delta$ $t<t_{\alpha}(n_1+n_2-2)$이면 귀무가설 기각
- 단측검정: 우측검정 H_0: $\mu_1-\mu_2=\delta$ vs H_1: $\mu_1-\mu_2>\delta$ $t>t_{1-\alpha}(n_1+n_2-2)$이면 귀무가설 기각

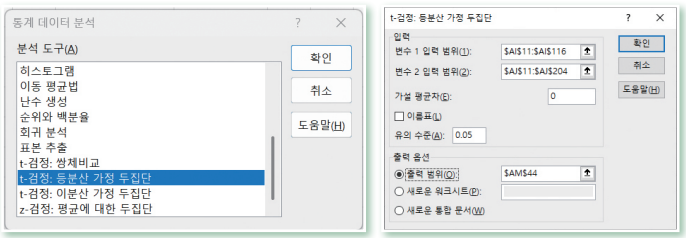
③ 모평균차이 $\mu_1-\mu_2$의 100(1-α)%신뢰구간

$$(\bar{x}-\bar{y}-t_{1-\alpha/2}(n_1+n_2-2)s\sqrt{(1/n_1+1/n_2)},\ \bar{x}-\bar{y}+t_{1-\alpha/2}(n_1+n_2-2)s\sqrt{(1/n_1+1/n_2)}$$

[실습] 남자와 여자의 BMI에 대한 평균 비교: 등분산 가정

위 F-검정을 통해 두 그룹의 분산이 같다는 것을 확인하였으므로, 같은 예제로 등분산 가정일 때의 두 집단의 모평균의 차이 검정을 위해 독립 이표본 t-검정을 수행해보자.

가. 위 예제와 같은 방법으로, 변수 1 입력 범위에는 남성의 BMI, 변수 2 입력 범위에는 여성의 BMI를 지정한다. 귀무가설은 “남자와 여자의 평균 BMI는 같다”이다. 유의수준은 5%으로 작성하였고, 출력 범위를 빈 셀로 지정하여 결과를 확인한다.



나. [$P(T\leq t)$ 단측 검정], [$P(T\leq t)$ 양측 검정] 모두 유의수준 0.05보다 작으므로, 귀무가설을 기각한다. 즉, 유의수준 5%에서 남자와 여자의 평균 BMI는 다르다고 할 수 있다.

t-검정: 등분산 가정 두 집단

	남성의 BMI	여성의 BMI
평균	24.33018868	23.58247423
분산	7.099460916	7.881763795
관측수	106	194
공동(Pooled) 분산	7.606120163	
가설 평균차	0	
자유도	298	
t 통계량	2.244643393	
P(T<=t) 단측 검정	0.012762287	
t 기각치 단측 검정	1.649982976	
P(T<=t) 양측 검정	0.025524574	
t 기각치 양측 검정	1.967956506	

(4) 두 독립 집단의 평균비교: 이분산 가정에서 독립 이표본 t-검정

다음으로 모분산이 동일하다고 가정하기 어려운 경우 두 집단의 표본분산을 사용하여 검정통계량과 신뢰구간을 아래와 같이 구할 수 있다. 이를 Satterthwaite의 근사라고 부른다.

① 검정통계량

검정통계량 t는 귀무가설하에서 근사적으로 자유도가 υ인 t-분포를 따른다. 이때 자유도 υ는 다음과 같다.

$$t=\frac{(\bar{x}-\bar{y})}{\sqrt{s_1^2/n_1+s_2^2/n_2}}\sim t(\upsilon),\ \mathrm{v}=\frac{(w_1+w_2)^2}{\frac{w_1^2}{n_1-1}+\frac{w_2^2}{n_2-1}},\ w_1=s_1^2/n_1,\ w_2=s_2^2/n_2$$

② 가설검정 결과

유의수준 α로 설정했을 때 가설에 따른 검정 결과는 다음과 같다.

- 양측검정: H_0: $\mu_1-\mu_2=\delta$ vs H_1: $\mu_1-\mu_2\neq\delta$ — $|t|>t_{1-\alpha/2}(\upsilon)$이면 귀무가설 기각
- 단측검정: 좌측검정 H_0: $\mu_1-\mu_2=\delta$ vs H_1: $\mu_1-\mu_2<\delta$ — $t<t_{\alpha}(\upsilon)$이면 귀무가설 기각
- 단측검정: 우측검정 H_0: $\mu_1-\mu_2=\delta$ vs H_1: $\mu_1-\mu_2>\delta$ — $t>t_{1-\alpha}(\upsilon)$이면 귀무가설 기각

③ 모평균차이 $\mu_1-\mu_2$의 100(1−α)%신뢰구간

$$(\bar{x}-\bar{y}-\mathrm{t}_{1-\alpha/2}(\mathrm{v})\sqrt{s_1^2/n_1+s_2^2/n_2},\ \bar{x}-\bar{y}-\mathrm{t}_{1-\alpha/2}(\mathrm{v})\sqrt{s_1^2/n_1+s_2^2/n_2})$$

[실습] 남자와 여자의 평균 체중 비교: 이분산 가정(ex_survey_data.xlsx)

이제 이분산 가정인 경우의 t-검정을 수행해 보자. 성별(남자, 여자)에 따라 체중 평균이 통계적으로 유의미한 차이가 있는지 확인하고자 한다.

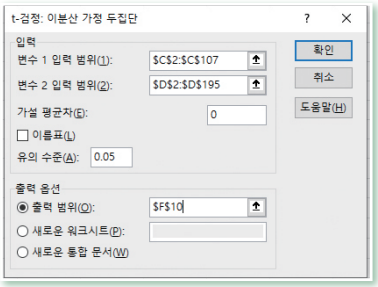
가. T_SEX 변수를 선택한 후, [홈] → [정렬 및 필터]를 통해 필터 기능을 사용한다. 성별이 1인 남성을 필터링 한 후, 체중의 수치를 복사하여, 새로운 시트에 데이터셋을 생성한다. 다음으로, 성별이 2인 여성을 필터링한 후, 체중의 수치를 복사하여 남성의 체중 오른쪽에 붙여넣는다.

나. 등분산 여부를 판단하려면 F-검정을 수행한다. 결과를 확인해본 결과, [$P(F\leq f)$ 단측 검정]의 값이 0.05보다 작으므로, 귀무가설을 기각한다. 즉, 두 집단의 분산이 다르므로 이분산인 것을 확인하였다.

F-검정: 분산에 대한 두 집단

	남성의 체중	여성의 체중
평균	69.54716981	57.74742268
분산	82.17394429	50.32447519
관측수	106	194
자유도	105	193
F 비	1.63288229	
P(F<=f) 단측 검정	0.001708017	
F 기각치: 단측 검정	1.31822705	

다. 변수 1 입력 범위에는 남성의 체중, 변수 2 입력 범위에는 여성의 체중을 지정한다. 귀무가설은 "남자와 여자의 평균 체중이 같다"이다. 유의수준은 0.05, 출력 범위를 빈 셀로 지정하여 결과를 확인한다.



라. [$P(T \le t)$ 단측 검정], [$P(T \le t)$ 양측 검정] 모두 유의수준 0.05보다 작으므로 귀무가설을 기각한다. 즉, 유의수준 5%에서 남자와 여자의 평균 체중은 다르다고 할 수 있다.

t-검정: 이분산 가정 두 집단

	남성의 체중	여성의 체중
평균	69.54716981	57.74742268
분산	82.17394429	50.32447519
관측수	106	194
가설 평균차	0	
자유도	176	
t 통계량	11.60059013	
P(T<=t) 단측 검정	8.92541E-24	
t 기각치 단측 검정	1.653557435	
P(T<=t) 양측 검정	1.78508E-23	
t 기각치 양측 검정	1.973534388	

3) 짝지은 표본에서 연속형 자료의 비교: 대응표본 t-검정

짝지은 표본은 서로 독립적인 두 모집단에서 추출된 표본이 아니라, 서로 관련이 있는 모집단에서 추출된 표본을 의미한다. 대표적인 예로는 동일한 개인을 대상으로 특정 처치 전후의 변화를 측정하는 교차설계가 있다. 예를 들어, 생활습관 개선 프로그램에 참가한 사람들의 체중 변화를 분석할 때, 프로그램 참가 전 체중과 참가 후 체중을 비교하면 이는 짝지은 표본에 해당한다. 또한, 같은 실험 조건에서 개인의 오른손과 왼손을 비교하거나, 짝지은 사례-대조 연구(matched case-control study)와 같이 환자와 유사한 특성을 가진 대조군을 1:1로 매칭하는 경우도 짝지은 표본의 예시가 된다.

짝지은 표본의 주요 장점은 동일한 개체나 유사한 집단 내에서 차이를 분석함으로써 처치나 조건의

효과를 보다 직접적으로 비교할 수 있다는 점이다. 이를 통해 집단 간 변동성을 줄이고, 분석의 신뢰도를 높일 수 있다. 그러나 짝지은 표본은 독립 표본과는 달리, 표본 간 상관성이 존재하기 때문에 이를 독립 표본으로 간주하여 분석하는 것은 부적절하다. 따라서 짝지은 표본에 적합한 통계적 분석 방법을 적용해야 한다.

이제 짝지은 표본에서 연속형 자료의 비교 검정 방법인 대응표본 t-검정(paired t-test)에 대해 알아보자. 다음과 같은 n명의 연구대상자가 있다고 하자.

번호	처리 전	처리 후	차이
1	x_1	y_1	d_1
2	x_2	y_2	d_1
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
n	x_n	y_n	d_n

여기서 $d_i=x_i-y_i(i=1, \cdots, n)$는 처리 전과 처리 후의 차이이다. 이때 x_i와 y_i는 서로 독립이 아니지만 연구대상자 간에는 독립이므로 d_i이 서로 독립이다.

(1) 가정

$d_1, d_2, \cdots, d_n$이 서로 독립이며, 모평균은 μ_d, 분산이 σ_d^2인 정규분포를 따른다.

(2) 검정통계량

검정통계량 t는 귀무가설하에서 근사적으로 자유도가 $n-1$인 t-분포를 따른다. 여기서 $\bar{d}$는 $d_i(i=1, \cdots, n)$의 표본평균이며, s_d는 표본표준편차이다.

$$t=\frac{\bar{d}-\delta}{(s_d/\sqrt{n})} \sim t=(n-1)$$

(3) 가설검정 결과

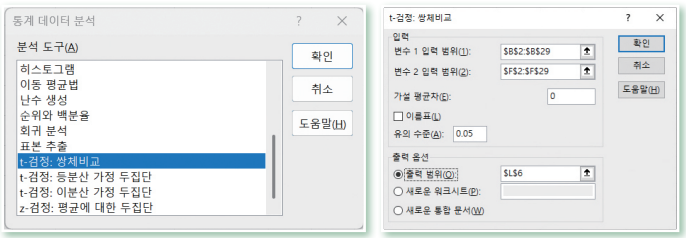
유의수준 α로 설정했을 때 가설에 따른 검정 결과는 다음과 같다.

- 양측검정: H_0: $\mu_d=\delta$ vs H_1: $\mu_d\neq\delta$ $|t|>t_{1-\alpha/2}(n-1)$이면 귀무가설 기각
- 단측검정: 좌측검정 H_0: $\mu_d=\delta$ vs H_1: $\mu_d<\delta$ $t<t_{\alpha}(n-1)$이면 귀무가설 기각
- 단측검정: 우측검정 H_0: $\mu_d=\delta$ vs H_1: $\mu_d>\delta$ $t>t_{1-\alpha}(n-1)$이면 귀무가설 기각

[실습] 대사증후군 관리 교육을 실시하기 전후 혈압의 비교

ex_pre_post_data.xlsx 데이터는 28명의 연구대상자를 대상으로 대사증후군 관리 교육을 실시하기 전후의 혈압을 측정한 자료이다. 수축기 혈압의 변화 여부를 분석하기 위해 대응표 t-검정(paired t-test)을 수행해보자. 이를 위해 Excel의 데이터 분석 도구에서 "t-검정: 쌍체 비교"기능을 사용하여 분석을 진행한다.

① t-검정: 쌍체비교를 선택해준 후, 변수 1에는 사전_수축기혈압, 변수 2에는 사후_수축기혈압을 지정한다. 귀무가설은 "대사증후군 관리 교육을 실시하기 전후의 혈압은 차이가 없다"이므로 가설 평균차에 "0"을 지정한다. 유의수준은 5%으로 작성하였고, 출력 범위를 빈 셀로 지정하여 결과를 확인한다.



② [$P(T\leq t)$ 단측 검정]은 유의수준 0.05보다 작으므로 귀무가설을 기각한다. 즉, 유의수준 5%에서 대사증후군 관리 교육을 실시하기 전후의 혈압은 차이가 있다고 할 수 있다.

t-검정: 쌍체 비교

	사전_수축기혈압	사후_수축기혈압
평균	135.8928571	125.4642857
분산	139.8029101	259.6653439
관측수	28	28
피어슨 상관 계수	-0.139689147	
가설 평균차	0	
자유도	27	
t 통계량	2.593578217	
P(T<=t) 단측 검정	0.007578746	
t 기각치 단측 검정	1.703288446	
P(T<=t) 양측 검정	0.015157491	
t 기각치 양측 검정	2.051830516	

3. 여러 집단의 평균 비교

1) 일원분산분석

지금까지는 두 집단 간 비교 문제를 다루었다. 그러나 연구에서는 서로 독립적인 다수의 집단에서 표본을 추출하여 비교하는 경우도 많다. 여러 집단을 비교하는 방법 중 하나로 각 집단을 짝지어 비교하는 방식이 있을 수 있다. 그러나 집단 수가 많아질수록 짝을 지어 비교해야 하는 횟수도 증가하며, 검정 횟수가 많아질수록 전체적인 제1종 오류(type I error)가 증가하는 문제가 발생한다. 따라서 여러 집단을 한 번에 전반적으로 검정(global test)할 수 있는 방법이 필요하며, 이를 위한 방법론이 분산분석(Analysis of Variance, ANOVA)이다.

분산분석은 총 변동(total variation)을 집단 간 변동(between-group variation)과 집단 내 변동(within-group variation)으로 나눈 후, 각 변동의 상대적인 크기를 계산하여 여러 집단의 모평균 차이를 검정하는 방법이다. 이는 집단의 표본평균 간 변동이 전체 표본 내 변동의 기댓값보다 더 크다면, 여러 집단이 동일한 모집단에서 추출되지 않았을 가능성이 높다는 개념에 기반한다.

분산분석은 결과에 영향을 미치는 요인(factor)의 개수에 따라, 관심 요인이 하나라면 일원분산분석(one-way ANOVA), 관심 요인이 둘이라면 이원분산분석(two-way ANOVA)이라고 한다. 예를 들어, 세 가지 지역(A도시, B도시, C도시)에서 평균 혈압 수준이 차이가 있는지를 분석하는 경우, 지역이라는 한 가지 요인만 고려하므로 일원분산분석을 수행한다. 반면, 흡연 여부와 운동방법이 혈압에 미치는 영향을 동시에 분석하는 경우, 흡연과 운동이라는 두 가지 요인을 고려하므로 이원분산분석을 수행한다.

이 교재에서는 많이 사용되는 일원분산분석에 대해 알아보자. 일원분산분석은 하나의 요인에 대한 여러 집단 간 평균 차이를 비교하는 방법으로, 역학 연구에서 질병 발생률, 혈압 수준, 체질량지수(BMI) 등의 차이를 지역, 연령대 등의 한 가지 요인에 따라 분석할 때 널리 활용된다.

k개의 수준을 가지는 요인이 있다고 하자. 예를 들어, 흡연 상태라는 범주형 변수가 "비흡연자", "과거 흡연자", "현재 흡연자"의 세 가지 수준으로 측정되는 경우, 흡연 상태는 3개의 수준을 가지는 요인이다.

(1) 가정

- k개의 모집단은 독립이며 정규분포를 따른다.
- 평균이 μ_i, 분산이 σ^2인 i번째 처리(i=1, ⋯, k) 모집단으로부터 n_i개의 표본을 랜덤하게 추출하였다.
- 각 처리에서 분산은 모두 σ^2으로 동일하다.

이때, 다음과 같은 모형을 설정한다. 여기서 ε_{ij}는 평균이 0, 분산이 σ^2인 정규분포를 따르는 오차항이다.

$$y_{ij}=\mu_i+\varepsilon_{ij},\ (i=1, 2, \cdots, k_{ij}=1, 2, \cdots, n_i)$$

일원분산분석의 귀무가설은 '각 집단의 모평균이 모두 같다'이며, 대립가설은 '각 집단의 모평균이 모두 같은 것은 아니다'로 설정된다.

$$H_0: \mu_1=\mu_2= \cdots =\mu_k \text{ vs } H_1: \mu_i\text{들이 모두 같지는 않다.}$$

분산분석에는 총제곱합(total sum of squares, SST)과 처리제곱합(treatment sum of squares, SSt), 그리고 오차제곱합(error sum of squares, SSE)이라는 개념이 필요하다. 먼저 총제곱합(SST)은 자료(표본)들이 전체 평균을 중심으로 어느 정도 변동이 있는지를 가늠하는 지표이며, 아래의 수식으로 정의된다.

$$SST=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2,\ df=n-1$$

처리제곱합(SSt)은 각 그룹(집단)의 평균이 전체 평균과 어느정도 차이가 있는지를 나타내는 지표이며, 그룹간 제곱합(between-group sum of squares)라는 용어로도 불리우며 아래와 같이 정의된다.

$$SSt=\sum_{i=1}^{k}n_i(\bar{y}_i-\bar{y})^2,\ df=k-1$$

오차제곱합(SSE)은 각 그룹(집단)내 자료와 해당 그룹의 평균의 변동성을 나타내는 지표로서 그룹내 제곱합(within-group sum of squares)로도 명명되며, 아래와 같이 정의된다.

$$SSE=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2,\ df=n-k$$

여기서 총제곱합과 처리제곱합, 오차제곱합 사이에는 다음과 같은 수식이 성립한다.

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2=\sum_{i=1}^{k}n_i(\bar{y}_i-\bar{y})^2+\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2$$

$$SST=SSt+SSE$$

즉, 전체변동은 처리에 따른 변동(처리제곱합)과 그 외의 모든 변동(오차제곱합)으로 분해할 수 있다.

일원분산분석에서 검정통계량은 제곱합을 자유도(degree of freedom, df)로 나눈 평균제곱합을 이용하여 다음과 같이 정의한다.

(2) 분산분석표

	제곱합	자유도	평균제곱합	F
처리	SSt	k-1	MSt=SSt/(k-1)	F=MSt / MSE
오차	SSE	n-k	MSE=SSE/(n-k)	
계	SST	n-1		

(3) 검정통계량

검정통계량 F는 집단 간 변동과 집단 내 변동의 비로 F값이 클수록 집단 간 변동이 집단 내 변동보다 큼을 의미한다. 이 F-통계량은 귀무가설하에서 분자와 분모의 자유도가 각각 k-1, n-k인 F-분포를 따른다.

$$F=\frac{SSt/(k-1)}{SSE/(n-k)}=\frac{MSt}{MSE}\sim F(k-1,\ n-k)$$

(4) 가설검정 결과

유의수준 α로 설정했을 때 $F>F_{1-\alpha}(k-1,\ n-k)$이면 귀무가설 H_0: $\mu_1=\mu_2=\cdots=\mu_k$을 기각한다.

[실습] 흡연 상태에 따라 혈압의 평균비교

Excel의 데이터 분석 도구에서는 "분산분석: 일원배치법" 기능을 사용하여 분산분석을 수행할 수 있다.

ex_survey_data.xlsx 데이터에서 비흡연, 과거흡연, 현재 흡연자의 수축기 혈압의 평균이 모두 같

은지 여부를 검정하여 보자. T_SMOKE는 흡연 상태를 나타내며, 1은 비흡연, 2는 과거 흡연, 3은 현재 흡연을 의미한다. T_SBP는 수축기 혈압을 나타내며, 본 예제에서 각 그룹 간 평균 차이를 검증하기 위해 사용되는 변수이다. 가설은 다음과 같다.

- H_0: 세 집단 간 수축기 혈압의 평균은 같다($\mu_1=\mu_2=\mu_3$).
- H_1: 세 집단 간 수축기 혈압의 평균은 모두 같지 않다.

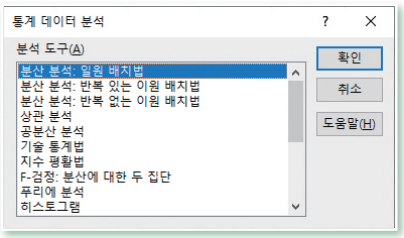
분산분석을 수행하기 위해 먼저, T_SMOKE 값을 기준으로 그룹화된 새로운 시트를 구성해보자.

① ex_survey 데이터에서 필터 기능을 사용하여 T_SMOKE 값(1, 2, 3)을 필터링하여 이때의 T_SBP 값을 복사한다.
② 새로운 시트(Sheet 2)에 필터링된 데이터를 그룹별로 붙여넣는다.
③ 각 T_SMOKE 값이 아래의 그림과 같이 하나의 열로 나뉘어 정리된 것을 확인할 수 있다.

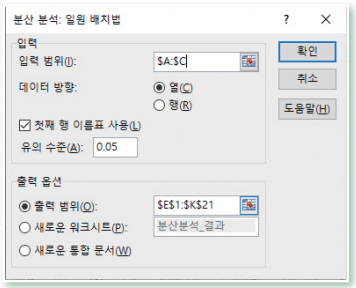
	A	B	C
1	T_SMOKE=1	T_SMOKE=2	T_SMOKE=3
2	134	110	141
3	130	110	141
4	121	118	160
5	124	110	115
6	110	126	140
7	134	123	122
8	107	140	125
9	110	120	128
10	142	108	99
11	110	120	109
12	117	140	108
13	130	124	118

이제 위의 세 그룹 간의 수축기 혈압의 평균 차이를 검증하기 위해 분산분석을 실행해보자.

① 데이터 → 데이터 분석 → 분산분석: 일원배치법을 선택한 후 확인을 클릭한다.



② [입력 범위]에 정리된 데이터의 전체 범위를 선택한 후, 첫째 행 이름표 사용을 클릭한다. 첫째 행 이름표 사용을 클릭하면 선택한 데이터의 첫 번째 행이 그룹 이름으로 사용된다. [유의수준]란에 원하는 유의수준을 입력하고 확인을 누른다.



③ 출력결과는 아래와 같다. 위의 요약표는 각 그룹에 대한 기초통계량을 요약한 표이며, 아래의 표는 분산분석 결과를 나타낸 표이다. F-통계량은 세 집단 간의 평균 차이를 나타내는 통계량이며, 결과표에서는 F 비로 표현되어 있다. F 기각치는 자유도 2와 297인 F-분포에서 상위 5% 임곗값을 나타낸다. 분산분석 결과에서 F-통계량이 F 기각치보다 작으므로, 유의수준 5% 하에서는 귀무가설을 기각할 수 없다. 따라서, 유의수준 5%에서 세 집단 간 수축기 혈압의 평균에 유의미한 차이가 없다고 결론지을 수 있다.

분산 분석: 일원 배치법

요약표

인자의 수준	관측수	합	평균	분산
T_SMOKE=1	208	25294	121.6057692	215.2931067
T_SMOKE=2	57	6839	119.9824561	164.2675439
T_SMOKE=3	35	4284	122.4	260.0117647

분산 분석

변동의 요인	제곱합	자유도	제곱 평균	F 비	P-값	F 기각치
처리	158.3144669	2	79.15723347	0.37552396	0.687255	3.026153
잔차	62605.05553	297	210.7914328			
계	62763.37	299				

6 범주형 자료의 비교

범주형 자료(categorical data)는 질적 자료(qualitative data)라고도 하며, 여러 개의 범주로 이루어진 자료를 의미한다. 자료가 연속적인 숫자가 아니라 특정한 범주나 수준으로 표현될 때 사용된다. 역학 연구에서는 질병 상태, 생활 습관, 위험 요인 등을 평가할 때 범주형 자료 형태로 측정하는 경우가 많다.

예를 들어, 감염병 역학조사에서 확진자의 질병 진행 상태는 “무증상”, “경증”, “중등도”, “중증”, “위중”과 같이 구분될 수 있다. 또한, 건강검진 데이터를 활용한 연구에서 비만도를 평가할 때 체질량지수(BMI)에 따라 “저체중”, “정상”, “과체중”, “비만”과 같은 범주로 나누어 분석할 수 있다. 이처럼 범주형

변수는 연속적인 숫자가 아닌 특정한 그룹이나 수준을 기준으로 데이터를 구분하고 해석하는 데 활용된다.

한편, 범주형 변수 중에서도 두 가지 수준만을 가지는 변수를 이항 변수라고 한다. 이항 변수는 "예/아니오", "있음/없음", "양성/음성"과 같이 두 가지 결과만을 가질 수 있다. 예를 들어, 백신 접종 여부를 나타내는 변수는 "접종함"과 "미접종"이라는 두 가지 값으로 표현될 수 있다. 이 장에서는 범주형 자료의 연관성을 설명하는 위험도에 대해 다루고, 연관성을 검정하는 방법인 카이제곱 검정(Chi-Square Test)에 대해 알아본다.

1. 연관성의 척도

두 범주형 변수의 연관성을 기술하기 위해서는 각 변수를 행과 열로 정리하고, 교차되는 칸에 각 범주에 해당하는 빈도수를 배치하는 분할표(contingency table)를 사용한다. 이 장에서는 분할표를 작성하는 방법을 살펴보고, 두 변수 간 연관성을 설명하는 주요 척도인 상대위험도(relative risk, RR)와 오즈비(odds ratio, OR)에 대해 알아본다.

1) 상대위험도

상대위험도(relative risk, RR)는 특정 요인과 결과 간의 연관성을 평가하는 척도로, 노출된 집단과 노출되지 않은 집단에서 질병 발생을 관찰하는 코호트연구(cohort study)나, 신약 또는 대조약을 할당한 후 결과를 관찰하는 임상시험(clinical trial)과 같은 연구 디자인에서 사용할 수 있다.

상대위험도는 다음과 같이 정의된다.

$$RR=\frac{\text{노출집단(실험군)에서의 위험도}}{\text{비노출집단(대조군)에서의 위험도}}=\frac{\text{노출집단(실험군)에서 사건이 발생할 확률}}{\text{비노출집단(대조군)에서 사건이 발생할 확률}}$$

상대위험도를 계산하기 위해, 요인과 결과가 각각 두 개의 범주로 구분된 분할표(contingency table)를 고려해보자. 이때, 행은 노출이 된 집단(E)과 노출되지 않은 집단($\overline{E}$)로 나누고, 열은 관심 질병의 발병 여부를 나타낸다. 전체 연구대상자 n_{++}명에 대해 다음과 같은 분할표를 작성할 수 있다.

노출(Exposure)	관심질병(Disease)		합계
	유(D)	무($\bar{D}$)	
예(E)	n_{11}	n_{12}	n_{1+}
아니오($\bar{E}$)	n_{21}	n_{22}	n_{2+}
합계	n_{+1}	n_{+2}	n

노출 집단(실험군)의 질병발생 위험도는 $risk_{\exp}=P(D|E)$이므로 $\frac{n_{11}}{n_{1+}}$으로 계산할 수 있고, 비노출 집단(대조군)의 질병발생 위험도=$risk_{\text{unexp}}=P(D|\bar{E})$이므로 $\frac{n_{21}}{n_{2+}}$으로 계산한다. 따라서 상대위험도 $RR=\frac{P(D|E)}{P(D|\bar{E})}$는 다음과 같이 추정할 수 있다.

$$\widehat{RR}=\frac{n_{11}/n_{1+}}{n_{21}/n_{2+}}$$

상대위험도가 1보다 크면 노출 집단의 질병 발생 위험이 비노출 집단보다 더 높다는 것을 의미한다. 반대로, 상대위험도가 1보다 작으면 노출 집단의 질병 발생 위험이 비노출 집단보다 낮다는 것을 의미하며, 이는 해당 노출이 질병 예방 효과를 가질 가능성이 있음을 시사한다.

상대위험도의 신뢰구간

상대위험도의 신뢰구간은 $\ln(RR)$이 근사적으로 정규분포를 따르는 특성을 이용하여 다음과 같이 구할 수 있다. 상대위험도의 신뢰구간에 대한 수리적 설명은 복잡하며, 이 책의 범위를 넘어가므로 생략한다.

$$\ln(RR)\text{의 } 100(1-\alpha)\% \text{ 신뢰구간: } \ln(\widehat{RR})\pm z_{\alpha/2}\sqrt{\frac{1}{n_{11}}-\frac{1}{n_{1+}}+\frac{1}{n_{21}}-\frac{1}{n_{2+}}}$$

$$RR\text{의 } 100(1-\alpha)\% \text{ 신뢰구간: } \widehat{RR}\times\exp\left(\pm z_{\alpha/2}\sqrt{\frac{1}{n_{11}}-\frac{1}{n_{1+}}+\frac{1}{n_{21}}-\frac{1}{n_{2+}}}\right)$$

상대위험도에 대한 신뢰구간의 해석은 신뢰구간이 1을 포함하지 않으면, 두 집단 간 상대위험도가 유의하다고 볼 수 있으며, 1을 포함하면 실험군과 대조군의 위험률은 차이가 없다고 해석할 수 있다.

2) 오즈비

오즈비(odds ratio, OR)는 두 집단에서 특정 사건이 발생할 오즈(odds)를 비교하는 척도로, 역학 및 의학 연구에서 위험 요인의 영향력을 평가하는 데 자주 사용된다. 오즈비는 주로 사례-대조 연구(case-control study)에서 사용되며, 상대위험도를 직접 계산할 수 없는 연구 디자인에서도 효과적으로 활용된다. 상대위험도 예제와 같이 다음과 같은 교차표를 고려해보자.

노출(Exposure)	관심질병(Disease)		합계
	유(D)	무($\overline{D}$)	
예(E)	n_{11}	n_{12}	n_{1+}
아니오($\overline{E}$)	n_{21}	n_{22}	n_{2+}
합계	n_{+1}	n_{+2}	n

오즈(odds)는 어떤 사건에 대해 그 사건이 발생하지 않은 확률에 대한 그 사건이 발생한 확률의 비로 정의된다. 즉, 노출 집단에서 질병에 대한 오즈는 $\text{odds}_{E}=P(D|E)/P(\overline{D}|E)$이며, 비노출 집단에서 질병에 대한 오즈=$\text{odds}_{\overline{E}}=P(D|\overline{E})/P(\overline{D}|\overline{E})$이다. 오즈비는 다음과 같이 정의된다.

$$\text{오즈비}(OR)=\frac{\text{노출집단에서의 오즈(odds)}}{\text{비노출집단에서의 오즈(odds)}}=\frac{P(D|E)/P(\overline{D}|E)}{P(D|\overline{E})/P(\overline{D}|\overline{E})}$$

여기서 노출 집단의 오즈는 $\frac{n_{11}\times n_{1+}}{n_{12}\times n_{1+}}$로 계산할 수 있고, 비노출 집단의 오즈는 $\frac{n_{21}\times n_{2+}}{n_{22}\times n_{2+}}$으로 계산한다. 따라서 오즈비는 다음과 같이 추정할 수 있다.

$$\widehat{OR}=\frac{n_{11}\times n_{22}}{n_{21}\times n_{12}}$$

오즈비가 1이라는 것은 노출 집단과 비노출 집단에서 특정 질병이 발생할 오즈가 동일하다는 의미이며, 이는 두 집단 간 노출과 질병 발생 간에 연관성이 없음을 나타낸다. 오즈비가 1보다 크면 노출 집단에서 질병이 발생할 오즈가 비노출 집단보다 크다는 의미이다. 즉, 노출이 질병 발생 위험을 증가시키는 요인으로 작용할 가능성이 있다. 예를 들어, 흡연자의 폐암 발생에 대한 오즈비가 2.5라면, 흡연자의 폐암 발생 오즈가 비흡연자의 2.5배라는 의미이다. 반대로, 오즈비가 1보다 작으면 노출 집단에서

질병이 발생할 오즈가 비노출 집단보다 작다는 의미이며, 이는 노출이 예방 효과를 가질 가능성이 있음을 나타낸다. 예를 들어, 백신 접종 후 감염에 대한 오즈비가 0.4라면, 백신 접종자가 미접종자보다 감염될 오즈가 60% 더 낮다는 의미이다.

오즈비의 신뢰구간

$$\ln(OR)\text{의 } 100(1-\alpha)\% \text{ 신뢰구간: } \ln(\widehat{OR}) \pm z_{\alpha/2}\sqrt{\frac{1}{n_{11}}+\frac{1}{n_{12}}+\frac{1}{n_{21}}+\frac{1}{n_{22}}}$$

$$OR\text{의 } 100(1-\alpha)\% \text{ 신뢰구간: } \widehat{OR} \times \exp\left(\pm z_{\alpha/2}\sqrt{\frac{1}{n_{11}}+\frac{1}{n_{12}}+\frac{1}{n_{21}}+\frac{1}{n_{22}}}\right)$$

오즈비의 신뢰구간도 상대위험도의 신뢰구간과 마찬가지로 1을 포함하면 실험군과 대조군의 오즈에 차이가 없다는 것을 의미한다.

[실습] 오즈비와 상대위험도

흡연 여부와 질병의 발생의 연관성을 기술하기 위해 오즈비와 상대위험도를 계산해보자. Excel의 데이터 분석 도구에는 오즈비(OR)와 상대위험도(RR)를 자동으로 계산하는 기능이 제공되지 않으므로, 수식을 직접 작성하여 계산해야 한다.

$$OR = \frac{(\text{질병이 있는 흡연자의 수}) \times (\text{질병이 없는 비흡연자의 수})}{(\text{질병이 있는 비흡연자의 수}) \times (\text{질병이 없는 흡연자의 수})}$$

$$RR = \frac{\text{흡연자의 질병발생률}}{\text{비흡연자의 질병발생률}}$$

$$\text{흡연자의 질병발생률} = \frac{\text{질병이 있는 흡연자의 수}}{\text{전체 흡연자의 수}}$$

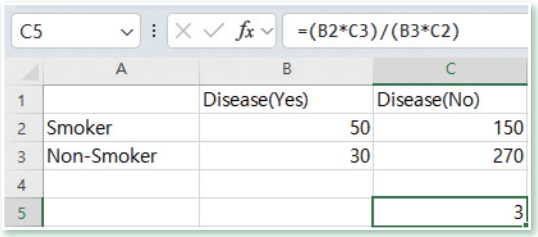
$$\text{비흡연자의 질병발생률} = \frac{\text{질병이 있는 비흡연자의 수}}{\text{전체 비흡연자의 수}}$$

위와 같은 수식을 이용하여 OR과 RR을 Excel에서 직접 계산해보자.

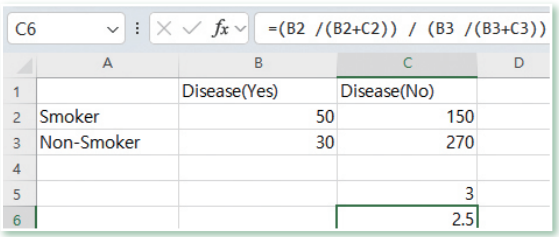
① 먼저, 위와 같은 데이터를 Excel에 직접 입력한다.

	A	B	C
1		Disease(Yes)	Disease(No)
2	Smoker	50	150
3	Non-Smoker	30	270

② OR 계산을 위해 C5 셀에 "=(B2*C3)/(B3*C2)"를 입력한다.



③ RR 계산을 위해 C6 셀에 "=(B2 /(B2+C2)) / (B3 /(B3+C3))"를 입력한다.



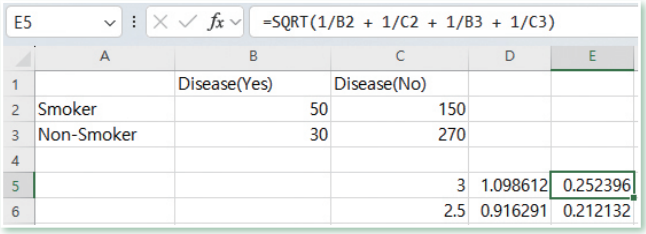
④ 계산된 셀의 값을 확인해보면, OR은 3으로 흡연자의 질병 발생 오즈가 비흡연자의 약 3배임을 알 수 있다. RR은 2.5로 흡연자의 질병 발생 위험이 비흡연자의 약 2.5배임을 알 수 있다.

[실습] 오즈비와 상대위험도의 신뢰구간

앞서 구한 오즈비 또는 상대위험도가 통계적으로 유의한지 평가하려면 추가적인 검정이 필요하다. 흔히 사용되는 방법은 OR과 RR의 신뢰구간을 계산하는 것이다. 신뢰구간은 추정치의 불확실성을 나타내며, 신뢰구간이 좁을수록 추정치를 더 신뢰할 수 있음을 의미한다.

이번 실습에서는 Excel을 이용하여 OR과 RR의 신뢰구간을 계산하는 방법을 알아본다. 이를 위해 이전에 사용했던 데이터를 그대로 활용하여 OR과 RR을 구하고, 신뢰구간을 추가로 계산할 것이다.

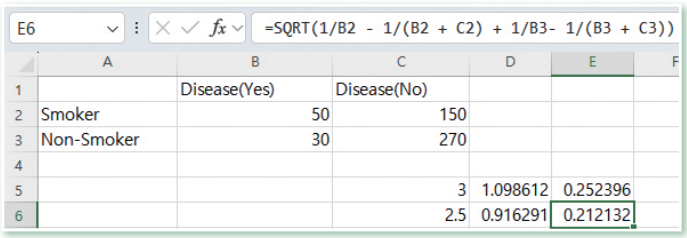
① 먼저, 앞서 구한 OR에 자연로그를 취한 ln(OR)을 구한다. 이을 위해 D5 셀에 "=LN(C5)"를 입력한다. 다음으로 ln(OR)의 표준오차를 계산하기 위해 E5 셀에 "=SQRT(1/B2+1/C2+1/B3+1/C3)"를 입력한다.



	A	B	C	D	E
1		Disease(Yes)	Disease(No)		
2	Smoker	50	150		
3	Non-Smoker	30	270		
4					
5			3	1.098612	0.252396
6			2.5	0.916291	0.212132

② OR의 신뢰구간 하한과 상한을 계산하기 위해 C8 셀에 "=EXP(D5-1.96*E5)"을 입력하고, D8 셀에 "=EXP(D5+1.96*E5)"를 입력한다.

③ RR의 신뢰구간을 계산하기 위해 먼저 자연로그를 취한 RR인 ln(RR)을 구한다. 이를 위해 D6 셀에 "=LN(C6)"를 입력한다, 다음으로 ln(RR)의 표준오차를 계산하기 위해 E6 셀에 "=SQRT(1/B2-1/(B2+C2)+1/B3-1/(B3+C3))"를 입력한다.



	A	B	C	D	E	F
1		Disease(Yes)	Disease(No)			
2	Smoker	50	150			
3	Non-Smoker	30	270			
4						
5			3	1.098612	0.252396	
6			2.5	0.916291	0.212132	

④ RR의 신뢰구간 하한과 상한을 계산하기 위해 C9 셀에 "=EXP(D6-1.96*E6)"을 입력하고, D9 셀에 "=EXP(D6+1.96*E6"를 입력한다.

⑤ 아래의 결과를 보면, OR은 95% 신뢰구간은 [1.829, 4.92]임을 알 수 있다. 이는 흡연자의 질병 발생 오즈가 비흡연자의 1.829배에서 4.92배 사이일 가능성이 있다. RR은 95% 신뢰구간은 [1.65, 3.79]임을 알 수 있다. 이는 흡연자의 질병 발생 위험이 비흡연자의 1.65배에서 3.79배 사이일 것이라고 95% 신뢰할 수 있음을 의미한다.

OR의 95% 신뢰구간	1.829268706	4.919999
RR의 95% 신뢰구간	1.649565541	3.788876

2. 연관성 검정

1) 카이제곱 검정

이 절에서는 두 범주형 변수의 연관성을 분석하는 방법으로 자주 사용되는 카이제곱 검정(Chi-Square Test)을 다룬다. 카이제곱 검정은 두 개 이상의 범주형 변수 간의 연관성이 유의미한지를 검정하는 방법이다. 이 검정은 표본의 빈도를 기반으로 실제 관찰값과 기대값 간의 차이를 평가하여 변수

간의 연관성을 분석하는 데 사용된다.

카이제곱 검정은 크게 두 가지로 구분된다.

- 동질성 검정(homogeneity test): 표본 수가 한 변수의 각 수준에 대해 미리 정해지는 경우로 이때 보고자 하는 것은 집단 간 각 수준별 결과 발생에 대한 분포가 동일한지를 검정하는 방법
  - H_0: 처리 1과 처리 2의 결과 발생 비율은 서로 같다.
- 독립성 검정(independence test): 전체 표본 수가 고정된 경우로 두 범주형 변수 간에 통계적으로 유의한 연관성이 있는지를 검정하는 방법
  - H_0: 처리의 종류와 결과 발생은 서로 독립이다.

두 가지 검정은 가설과 유도 과정은 다르지만, 결과적으로 동일한 검정 방법을 제공한다. 따라서, 이를 포괄하여 카이제곱 검정이라고 부르겠다. 다음과 같은 2×2 분할표를 고려하자.

	결과1	결과2	합
처리 1	n_{11}	n_{12}	n_{1+}
처리 2	n_{21}	n_{22}	n_{2+}
합	n_{+1}	n_{+2}	n

만약 처리의 종류와 결과 변수가 서로 독립이라면, 다음과 같은 빈도를 기대할 수 있다.

	결과1	결과2
그룹 1	$E_{11}=n_{1+}n_{+1}/n$	$E_{12}=n_{1+}n_{+2}/n$
그룹 2	$E_{21}=n_{2+}n_{+1}/n$	$E_{22}=n_{2+}n_{+2}/n$

두 변수가 서로 독립이라는 귀무가설이 참이라면 기대빈도와 관찰빈도는 서로 유사할 것이다. 그러나 기대빈도와 관찰빈도 간의 차이가 크게 나타난다면, 이는 처리와 결과가 서로 독립이라는 귀무가설에 반하는 증거가 될 수 있다. 이에 근거한 카이제곱 검정통계량(Pearson's chi-square statistic)은 다음과 같이 정의된다.

$$X^2=\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(n_{ij}-E_{ij})^2}{E_{ij}}\sim x^2(1),\ where\ E_{ij}=\frac{n_i+n_{+j}}{n}$$

X^2 통계량은 자유도가 1인 X^2-분포를 따른다.

유의수준을 α로 설정했을 때, $X^2>X^2_{1-\alpha}(1)$이면 귀무가설을 기각한다.

이제 $I\times J$ 분할표에서 카이제곱 검정을 수행해보자. 이때 검정통계량은 다음과 같이 정의된다.

$$X^2=\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij}-E_{ij})^2}{E_{ij}}\sim x^2((I-1)(J-1)),\ where\ E_{ij}=\frac{n_i+n_{+j}}{n}$$

X^2 통계량은 자유도가 $(I-1)(J-1)$인 x^2-분포를 따른다.

유의수준을 α로 설정했을 때, $X^2>x^2_{1-\alpha}((I-1)(J-1))$이면 귀무가설을 기각한다.

[실습] 흡연 여부와 질병의 연관성 검정

흡연 여부와 질병이 독립인지를 검정하기 위해 Excel을 이용하여 카이제곱 검정을 해보자. 검정을 위한 귀무가설(H_0)로 "흡연 여부와 질병은 서로 독립이다"라고 설정한다.

① 먼저 관측값의 행과 열의 합계를 계산한다. 이를 통해 기대도수를 계산하는 데 필요한 열 합과 행 합, 그리고 전체 합계를 구할 수 있다. Excel에서 합계를 구하려면 D2 셀을 클릭한 뒤 상단 메뉴의 "수식" 탭에서 "자동 합계" 버튼을 누른다. 나머지 행과 열의 합계도 동일한 방법으로 계산하여, 모든 합계를 구한다.

	A	B	C	D
1		Disease(Yes)	Disease(No)	계
2	Smoker	50	150	200
3	Non-Smoker	30	270	300
4	계	80	420	500

② 기대값은 각 셀의 열합과 행합을 곱한 후 전체 합으로 나누어 계산한다. 예를 들어, Disease(Yes)이면서 Smoker인 셀의 기대값은 Disease(Yes)의 열합(총합)과 Smoker의 행합을 곱한 뒤 전체 합으로 나눈 값이다. 이를 Excel로 계산하기 위해 B7 셀에 "=(B$4*$D2)/$D$4"을 입력한다. 마찬가지로 나머지 기대도수를 구하면 된다. 이때, 앞의 장에서 실습했던 절대참조("$")를 사용함으로써, 수식을 다른 셀로 드래그하여 복사할 때 참조가 올바르게 유지된다. 나머지 기대값도 동일하게 계산하며, 기대값 범위(B7:C8)의 행과 열 합계도 자동 합계를 사용해 계산한다.

	기대치		
	Disease(Yes)	Disease(No)	계
Smoker	32	168	200
Non-Smoker	48	252	300
계	80	420	500

③ 기대값과 관측값을 비교하여 카이제곱 검정을 수행한다. 먼저 A11 셀에 “p-value”, A12 셀에 “검정결과”를 입력한다. 관측값 범위는 B2:C3, 기대값 범위는 B7:C8이므로, p-value를 구하기 위해 B11 셀에 “=CHISQ.TEST(B2:C3, B7:C8)”을 입력한다. 이 수식은 관측도수와 기대도수를 비교하여 카이제곱 검정의 p-value를 반환한다. 이후, p-value가 유의수준(0.05)보다 작으면 귀무가설을 기각하도록 검정 결과를 표시한다. 이를 위해 B12 셀에 “=IF(B11<0.05, “귀무가설 기각”, “귀무가설 기각X”)”를 입력한다.

p-value	=CHISQ.TEST(B2:C3, B7:C8)
검정결과	=IF(B11<0.05, "귀무가설 기각", "귀무가설 기각x")

④ 이번에는 카이제곱 통계량과 카이제곱 분포에서의 임곗값을 구해보자. 먼저 A14, A15, A16, A17 셀에 각각 “자유도”, “유의수준”, “검정통계량”, “임곗값”을 입력한다. 본 실습에서 데이터는 2×2 행렬이므로 자유도는 (행의개수−1)×(열의개수−1)=1이다. 따라서 B14 셀에 1을 입력한다. 유의수준은 5%로 설정하고, B15 셀에 0.05를 입력한다. 카이제곱 검정통계량은 관측값과 기대값의 차이를 기반으로 계산된다. Excel에서는 SUMPRODUCT 함수를 사용해 카이제곱 통계량을 계산할 수 있다. 이를 위해 B16 셀에 “=SUMPRODUCT(B2:C3-B7:C8, B2:C3-B7:C8, (B7:C8)^(-1))”를 입력한다. 마지막으로 임곗값을 얻기 위해 B17 셀에는 “=CHISQ.INV.RT(B15,B14)”를 입력한다.

자유도	1
유의수준	0.05
검정통계량	20.08928571
임계값	3.841458821

⑤ 위의 계산을 통해 카이제곱 검정통계량은 20.089, 임곗값은 3.841임을 확인할 수 있다. 검정통계량(20.089)이 임곗값(3.841)보다 크므로, 귀무가설을 기각한다. 이는 유의수준 5%에서 흡연 여부와 질병의 유무는 서로 연관이 있음을 의미한다.

2) 피셔의 정확 검정

피셔의 정확검정은 두 범주형 변수 간의 연관성을 검정할 때 사용되는 통계적 방법으로, 기대도수가 작은 경우(기대도수가 5 미만인 셀이 전체 셀의 20% 이상일 때)에도 초기하분포에 기반하여 정확한 p-value를 계산할 수 있어, 표본크기가 작을 때 카이제곱 검정보다 더 적절하다.

(1) 가설

H_0: 변수1과 변수2는 서로 독립이다.

H_1: 변수1과 변수2는 서로 독립이 아니다.

		변수2		합
		Y1	Y2	
변수1	X1	a	b	a+b
	X2	c	d	c+d
	합	a+c	b+d	a+b+c+d=N

(2) 검정통계량

분할표에서 행의 합과 열의 합이 각각 고정되어있을 때, 변수1은 X1이고 변수2는 Y1일 때의 빈도를 확률변수라고 할 때 확률변수는 초기하분포를 따른다. 이에 기반하여 위 분할표가 나올 확률은 다음과 같이 계산할 수 있다.

$$P(X=a)=\frac{\left(\frac{a+c}{a}\right)\left(\frac{b+d}{b}\right)}{\left(\frac{N}{a+b}\right)}=\frac{(a+b)!(c+d)!(a+c)!(b+d)!}{N!a!b!c!d!}$$

이때 p-value는 관찰된 칸의 빈도보다 더 극단적인 결과가 나올 확률을 모두 더해 계산한다. 양측검정의 p-value는 극단적인 결과가 나올 확률(단측검정의 p-value)에 2배를 곱하여 계산한다. p-value와 유의수준 α를 비교하여, p-value$<\alpha$인 경우 변수 1과 변수 2가 독립이라는 귀무가설을 기각한다.

3) 짝지은 표본의 연관성 검정: 맥니마 검정

맥니마 검정(McNemar test)은 짝지은 표본에서 두 범주형 변수 간의 변화 또는 차이를 검정하는 방법이다. 주로 같은 대상에게 두 번 측정한 결과나 짝지어진 관측값을 비교할 때 수행한다. 다음의 2×2 분할표를 고려하자. 같은 대상자에게서 사전과 사후를 측정하였다고 하자.

사전	사후		합
	예	아니오	
예	n_{11}	n_{12}	n_{1+}
아니오	n_{21}	n_{22}	n_{2+}
합	n_{+1}	n_{+2}	n

(1) 가설

H_0: 사전과 사후에 반응률이 서로 같다($p_1=p_2$).

H_1: 사전과 사후에 반응률이 서로 같지 않다($p_1 \neq p_2$).

여기서 사전의 반응율은 p_1, 사후의 반응율은 p_2이다. 이들의 추정량은 다음과 같다.

$$\hat{p_1}=\frac{n_{1+}}{n},\ \hat{p_2}=\frac{n_{+1}}{n}$$

(2) 검정통계량

맥니마 검정의 검정통계량은 사전과 사후의 불일치 쌍인 n_{12}와 n_{21}을 사용하여 다음과 같이 정의된다. 검정통계량 Q_M은 귀무가설하에서 자유도가 1인 카이제곱분포를 따른다.

$$Q_M=\frac{(n_{12}-n_{21})^2}{n_{12}+n_{21}} \sim \chi^2(1)$$

유의수준을 α로 설정했을 때, 검정통계량 $Q_M > \chi^2_{1-\alpha}(1)$이면 귀무가설을 기각한다.

7 상관분석의 이해 및 실습

1. 상관분석

상관분석(correlation analysis)은 두 변수 간의 관계를 정량적으로 측정하는 분석 방법이다. 특정 변수의 값이 변할 때 다른 변수의 값도 함께 변하는 경향이 있는지를 분석하는 데 사용된다. 예를 들어, 신체활동 수준과 BMI 간의 관계를 연구한다고 가정해보자. 신체활동이 증가할수록 BMI가 감소하는 경향이 있다면, 두 변수 간에는 음의 상관관계(negative correlation)가 존재한다고 할 수 있다. 상관분석은 역학, 사회과학, 경제학, 심리학 등 다양한 분야에서 변수 간의 관계를 분석하는 데 활용된다.

두 변수 간의 상관 정도를 수치로 나타내는 값을 상관계수(correlation coefficient)라고 하며, 가장 널리 사용되는 상관계수는 피어슨 상관계수(Pearson correlation coefficient)로 다음과 같이 정의된다.

$$r=\frac{\Sigma(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\Sigma(x_i-\overline{x})^2}\sqrt{\Sigma(y_i-\overline{y})^2}}$$

여기서 X_i와 Y_i는 각각 변수 X와 Y의 관측값이며, $\overline{X}$와 $\overline{Y}$는 각각 변수 X와 Y의 표본평균이다. 상관계수 r은 -1에서 1사이이며, 그림 1-1-10과 같이 r=1이면 완전한 양의 상관관계이고 r=-1이면 완전한 음의 상관관계이며, 부호에 따라 두 변수가 양의 관계인지 음의 관계인지 알려준다. r이 -1이나 1에 가까울수록 선형적 관련성의 정도는 더 커진다. 반대로 r이 0에 가깝다는 것은 선형적 관련성의 정도가 없다는 것인데 이것은 X와 Y가 연관성이 없다는 것과 다르다.

그림 1-1-11의 (a)에서는 X와 Y가 강한 상관관계를 가지는 것처럼 보이는데 실제로 r=0이다. 이는 X와 Y가 이차곡선 형태의 비선형적 관계를 가지기 때문이다. 상관계수 r은 선형적 관계를 측정하는 척도이므로, 이러한 비선형적 관계를 설명할 수 없다. (b)에서는 두 개의 이상치(outlier)가 자료의 상관관계에 큰 영향을 미쳐, 실제로는 X와 Y에는 상관관계가 거의 없음에도 불구하고 상관계수 r이 큰 값으로 계산될 수 있다. 이러한 경우, 이상치로 인해 상관계수가 왜곡될 가능성이 높으므로 상관계수 r만을 기반으로 두 변수 간의 관계를 해석하는 것은 부적절하다. 따라서 비선형적 관계나 이상치가 존재하는 경우에는 상관계수 해석에 주의해야 한다. 또한, 상관분석에서 두 변수 간의 상관관계가 존재한다고 해서 반드시 인과관계(causal relationship)가 성립하는 것은 아니므로 단순히 상관분석의 결과를 인과관계로 해석하지 않도록 주의해야 한다.

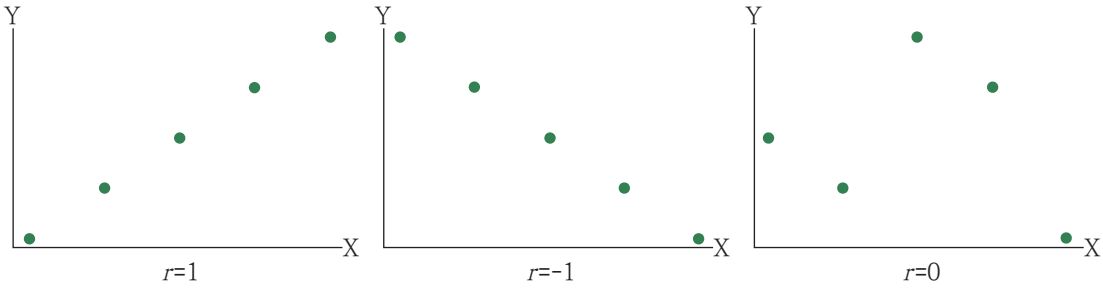


그림 1-1-10. *r*값의 크기에 대응하는 산점도

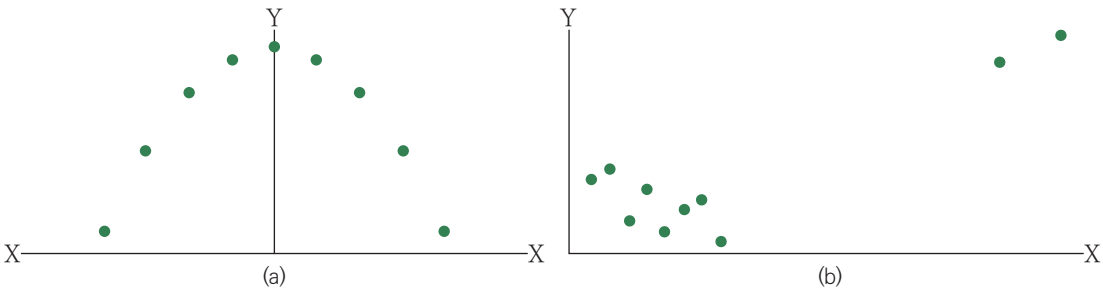


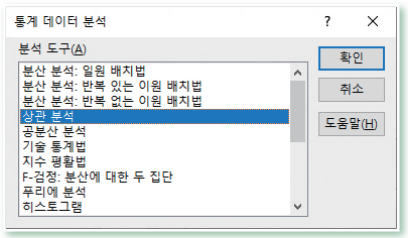
그림 1-1-11. 상관계수로 상관관계를 해석하는데 부적절한 산점도

2. 상관분석 실습

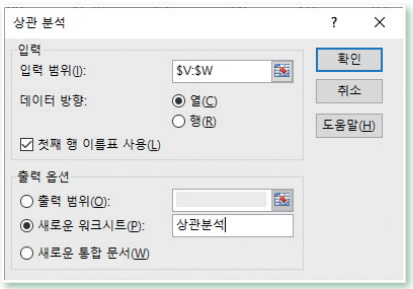
상관분석은 두 변수 간의 선형 관계의 강도와 방향을 측정하는 통계 기법이다. 상관계수는 -1과 1사이의 값을 가지며, 0에 가까울수록 상관관계가 약함을 나타낸다. Excel의 데이터 분석 기능에서 상관분석을 진행할 수 있다.

이번 실습에서는 Excel을 사용하여 ex_survey_data.xlsx 데이터에서의 체중(T_WEIGHT)과 BMI(T_BMI)의 상관계수를 계산하고, 상관계수의 유의성을 검정하는 방법을 알아보자.

① 데이터 → 데이터 분석 → 상관 분석을 선택한다.



② [입력 범위]란에 "$V:$W"을 입력하고 '첫째 행 이름표 사용'을 선택한다. 본 예제에서는 [출력 옵션]은 새로운 워크시트에 출력하도록 설정해주었다.



③ 출력결과는 다음과 같이 상관계수 행렬 형태로 나타난다. B2 셀은 T_WEIGHT와 T_WEIGHT 간의 상관계수를 나타내며, 동일 변수 간의 상관계수는 항상 1로 표시된다. C3도 동일하게 자신과의 상관계수를 나타내므로 값이 1이다. 본 실습에서 관심은 몸무게와 BMI의 상관계수이다. B3 셀을 확인해보면 약 0.759인 것으로 나타나며, 이는 두 변수 간에 강한 양의 상관관계가 있음을 의미한다.

	A	B	C
1		T_WEIGHT	T_BMI
2	T_WEIGHT	1	
3	T_BMI	0.758903	1

④ 피어슨 상관계수 검정을 진행해보자. 피어슨 상관계수 검정은 표본 상관계수를 이용하여 두 변수 간의 상관관계가 통계적으로 유의한지 평가한다. 이 검정의 귀무가설(H_0)은 "두 변수 간에 상관관계가 없다"로 설정할 수 있다. 검정통계량을 계산하기 위해서 B6 셀에 "=B3*SQRT((B5-2)/(1-B3^2))"를 입력한다. 또한 유의확률 P-값을 계산하기 위해 B7 셀에 "=T.DIST.2T(ABS(B6), B5-2)"를 입력한다.

	A	B	C
1		T_WEIGHT	T_BMI
2	T_WEIGHT	1	
3	T_BMI	0.758902505	1
4			
5	표본수	300	
6	검정통계량	20.11763322	
7	p-값	1.8662E-57	

⑤ 위와 같은 결과를 얻게 된다. P-값이 매우 작으므로(<0.05) 유의수준 5%에서 몸무게와 BMI 간의 상관계수가 통계적으로 유의미하다고 결론지을 수 있다.

8 회귀분석의 이해 및 실습

1. 회귀분석

회귀분석(regression analysis)은 변수들 간의 관계를 분석하고 예측 모델을 구축하는 통계적 방법이다. 주어진 독립변수(설명변수)를 이용하여 종속변수(반응변수)의 변화를 설명하거나 예측하는 데 사용된다. 앞서 독립변수 X와 종속변수 Y 간의 관계를 산점도를 이용하여 시각화하였다. 만약 자료가 근사적으로 직선의 형태를 보인다면, 두 변수는 선형적 관계를 가진다고 판단할 수 있다. 이러한 두 변수 간의 관계를 설명하기 위해 회귀분석(linear regression analysis)을 수행한다. 예를 들어, 체질량지수(BMI)와 혈압 간의 관계를 분석하려 한다면, 회귀분석을 통해 BMI가 혈압에 미치는 영향을 수량적으로 평가할 수 있다. 회귀분석은 다양한 연구 분야에서 널리 활용되며, 역학 연구에서는 건강 지표와 생활 습관, 질병 발생 위험 요인 간의 관계를 분석하는 데 자주 사용된다.

2. 단순회귀분석

회귀분석은 독립변수의 개수에 따라 단순회귀분석(simple regression)과 다중회귀분석(multiple regression)으로 구분된다. 이 절에서는 단순회귀분석에 대해 알아본다.

n개의 표본에서 독립변수와 종속변수의 쌍 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$의 데이터를 수집하였다고 하자. 단순회귀모형은 하나의 독립변수 X를 이용하여 종속변수 Y설명하기 위해 다음과 같은 회귀모형을 설정한다.

$$y=\beta_0+\beta_1 x_i+\varepsilon_i,\ i=1,\ \cdots,\ n$$

여기서 β_0와 β_1은 회귀계수(regression coefficient)라고 부르며, ε_i은 오차항(error term)으로 회귀직선으로 설명할 수 없는 예측 불가능한 변동을 의미한다. ε_i는 평균이 0이고 분산이 σ^2인 정규분포를 따르며, ε_i들은 서로 독립이라고 가정한다.

이제 자료로부터 회귀계수를 추정해보자. 우리는 표본을 사용하여 회귀분석을 실시한다. 실제 모수 β_0와 β_1을 표본자료로부터 추정해야 한다. 추정방법으로는 주로 최소제곱법(method of least squares)을 사용한다. 최소제곱법은 오차제곱합을 최소로 하는 β_0와 β_1을 찾는 방법이다. 이를 위해 오

차들의 제곱합을 β_0와 β_1에 대해 편미분하여 추정치 $\hat{\beta}_1$와 $\hat{\beta}_0$을 구한다.

1) 회귀계수에 대한 추론

회귀계수의 유의성을 평가하기 위해 가설검정을 해볼 수 있다. 만약 회귀계수가 0이라면, 독립변수 X가 어떤 값을 취하더라도 종속변수 Y에 영향을 미치지 않음을 의미한다. 즉, 두 변수 간에 선형 관계가 없음을 의미한다. 회귀계수에 대한 가설은 다음과 같이 설정할 수 있다.

- 양측검정: H_0: β=0 vs H_1: $\beta \neq 0$
- 단측검정: 좌측검정 H_0: β=0 vs H_1: $\beta < 0$
- 단측검정: 우측검정 H_0: β=0 vs H_1: $\beta > 0$

만약 H_0: β=0이라는 귀무가설이 참이라면 독립변수 X와 종속변수 Y는 선형 관계가 없다.

(1) 검정통계량

검정통계량 t는 H_0: β=0이라는 귀무가설 하에서 자유도가 n-2인 t-분포를 따른다.

$$t=\frac{\hat{\beta}}{SE(\hat{\beta})} \sim t(n-2)$$

여기서 $SE(\hat{\beta})$는 $\hat{\beta}$의 표준오차이다.

(2) 가설검정의 결과

유의수준을 α로 설정했을 때 가설에 대한 검정 결과는 다음과 같다.

- 양측검정: H_0: β=0 vs H_1: $\beta \neq 0$ — $|t| > t_{1-\alpha/2}(n-2)$이면 귀무가설 기각
- 단측검정: 좌측검정 H_0: β=0 vs H_1: $\beta < 0$ — $t < t_{\alpha}(n-2)$이면 귀무가설 기각
- 단측검정: 우측검정 H_0: β=0 vs H_1: $\beta > 0$ — $t > t_{1-\alpha}(n-2)$이면 귀무가설 기각

(3) 회귀계수 β에 대한 100(1-α)% 신뢰구간

$$(\hat{\beta}-t_{1-\alpha/2}(n-2)\times SE(\hat{\beta}),\ \hat{\beta}+t_{1-\alpha/2}(n-2)\times SE(\hat{\beta}))$$

여기서 $t_{1-\alpha/2}(n-2)$는 자유도가 $n-2$인 t-분포에서 하위 $1-\alpha/2$의 확률을 갖는 임곗값이다.

2) 결정계수

회귀모형이 데이터를 얼마나 잘 설명하는지를 평가하는 방법 중 하나는 결정계수(R^2)를 이용하여 회귀식이 전체 변동 중 어느 정도를 설명하는지 확인하는 것이다. 결정계수는 다음과 같이 총제곱합(total sum of squares, SST)과 회귀식으로 설명되는 변동인 회귀제곱합(regression sum of squares, SSR)의 비로 정의된다. 여기서, 총제곱합은 종속변수의 총 변동량이며, 회귀식으로 설명되는 변동인 회귀제곱합(SSR)과 회귀식으로 설명되지 않는 오차제곱합(SSE)으로 분해된다.

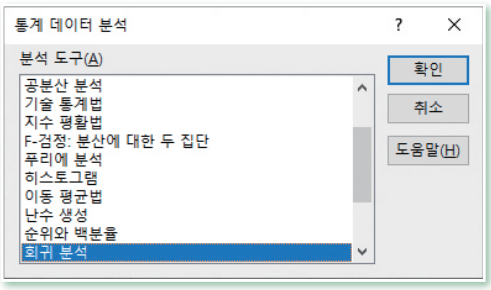
$$R^2=\frac{SSR}{SST}=\frac{1-SSE}{SST}$$

R^2값은 1에 가까울수록 데이터의 변동을 회귀식으로 설명되는 부분이 크다는 의미이며, R^2값이 0에 가깝다면 거의 설명되는 부분이 없다는 것을 의미한다.

[실습] 단순회귀분석(ex_survey_data.xlsx)

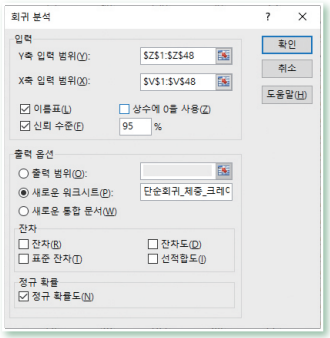
단순회귀분석은 하나의 독립변수와 하나의 종속변수 간의 관계를 분석하는 통계 방법이다. Excel을 통해 단순회귀분석을 수행할 수 있다. Creatinine은 근육에서 생성되는 노폐물로, 대사 과정에서 생성된 후 신장을 통해 배설된다. 체중이 증가할수록 혈중 Creatinine의 농도도 같이 증가할 가능성이 있다. 따라서 이번 실습에서는 40~43세의 환자를 대상으로 체중(T_WEIGHT)이 혈액 Creatinine (T_CREATINE) 농도에 미치는 영향을 단순회귀분석을 통해 분석해보자.

① 데이터 → 데이터 분석 → 회귀분석을 선택한다.



② [Y축 입력 범위]란에는 반응변수가 들어있는 "$Z$1:$Z$48"을, [X축 입력 범위]란에는 설명변수

의 범위 "$V$1:$V$48"을 입력한다. 데이터 범위에 변수명이 포함되어있으므로 '이름표'를 체크한다. '신뢰수준'은 결과에서 신뢰 구간을 계산할 때 이용한다. '상수에 0을 사용'은 절편을 강제로 0으로 설정하는 것이므로, 일반적으로는 체크하지 않는다. 본 예제에서는 [정규 확률]에서 '정규 확률도'를 선택하여 잔차가 정규 분포를 따르고 있는지 확인하기 위한 그림을 그리도록 한다. 마지막으로 확인을 누른다.



③ 주요 출력결과는 다음과 같다.

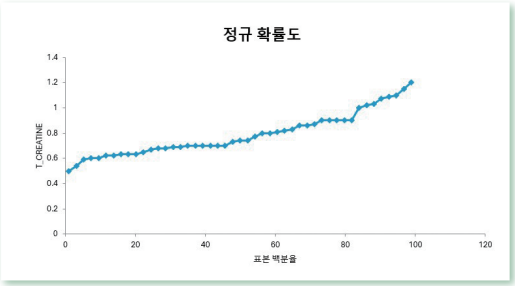
먼저, 회귀분석 통계량에서 다중 상관계수란 종속 변수와 독립 변수 간의 상관관계를 나타내는 통계량이다. 값이 0에 가까울수록 상관관계가 약하며, 1에 가까울수록 강한 상관관계를 의미한다. 여기서는 상관계수가 0.5119로 중간 정도의 양의 상관관계가 있음을 의미한다. 또한, 결정계수는 독립 변수가 종속 변수의 변동을 얼마나 잘 설명하는지 나타내는 계수이다. 결정계수는 0.262로, 독립 변수가 종속 변수의 약 26.2%를 설명한다는 것을 알 수 있다.

다음으로 분산분석에서는 회귀모형이 유의한지를 F-검정한다. F 통계량이 클수록 독립 변수가 종속 변수에 미치는 영향이 크다는 것을 의미하는데, F 통계량이 15.976이며 유의한 F 값(p-value)은 0.05보다 작으므로 귀무가설을 기각한다. 즉, 체중이 혈액 Creatinine 농도에 통계적으로 유의미한 영향을 미친다는 것을 알 수 있다.

아래의 출력 결과의 마지막 표에서는 회귀계수에 대한 검정통계량과 신뢰구간을 제공한다. T_WEIGHT의 계수는 0.00702으로 체중이 1단위 증가할 때, 혈액 Creatinine 농도는 평균적으로 0.00702 mg/dL 증가하는 것으로 해석할 수 있다. p-value는 유의수준 0.05보다 작으므로, 유의수준 5%에서 체중이 혈액 Creatinine 농도에 유의미한 영향을 미친다는 것을 확인할 수 있다. 또한, 95% 신뢰구간은 0.003483에서 0.010558이다.

	A	B	C	D	E	F	G	H	I
1	요약 출력								
2									
3	회귀분석 통계량								
4	다중 상관계수	0.511868453							
5	결정계수	0.262009313							
6	조정된 결정계수	0.24560952							
7	표준 오차	0.147437516							
8	관측수	47							
9									
10	분산 분석								
11		자유도	제곱합	제곱 평균	F 비	유의한 F			
12	회귀	1	0.347292	0.347292	15.97638	0.000235			
13	잔차	45	0.978202	0.021738					
14	계	46	1.325494						
15									
16		계수	표준 오차	t 통계량	P-값	하위 95%	상위 95%	하위 95.0%	상위 95.0%
17	Y 절편	0.35338263	0.110705	3.192104	0.002577	0.130411	0.576354	0.130411	0.576354
18	T_WEIGHT	0.007020308	0.001756	3.997046	0.000235	0.003483	0.010558	0.003483	0.010558

④ 화면의 오른쪽에는 정규 확률도가 표시된다. 정규 확률도란 데이터가 정규 분포를 따르는지 시각적으로 평가하기 위해 사용되는 도구이다. 데이터가 정규 분포를 따를 경우, 그래프는 대각선에 가까운 직선 형태를 나타낸다. 그래프에서 데이터 점들이 대체로 대각선에 가까운 직선 위에 위치해 있음을 알 수 있다.



3. 다중회귀분석

단순회귀분석은 하나의 독립변수가 종속변수에 미치는 영향을 분석하는 데 사용된다. 하지만 현실에서는 종종 여러 변수들이 동시에 종속변수에 영향을 미친다. 이를 효과적으로 분석하기 위해 다중회귀분석(multiple regression analysis)을 사용한다. i번째 개체에서 종속변수 y_i와 p개의 독립변수 x_{1i}, …, x_{pi}가 있을 때 다중회귀모형은 다음과 같이 표현된다.

$$y_i=\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+ \cdots +\beta_p x_{pi}+\varepsilon_i,\ i=1, \cdots, n$$

여기서 오차항 ε_i는 평균이 0, 분산이 σ^2인 정규분포를 따르면 서로 독립으로 가정한다. 다중회귀분석에서 β_j는 회귀계수로 단순회귀분석에서의 회귀계수와 다소 의미가 다르다. 단순회귀분석에서 회귀

계수는 한 독립변수가 종속변수에 미치는 영향을 직접 해석할 수 있지만, 다중회귀분석에서는 다른 독립변수들의 영향을 통제한 상태에서 특정 독립변수가 종속변수에 미치는 순수한 효과를 나타낸다는 점에서 차이가 있다.

1) 회귀계수에 대한 추론

다중회귀에서도 단순회귀와 마찬가지로 최소제곱법으로 회귀계수 β_0, β_1, … ,β_p를 추정한다. 회귀계수 β_j에 대해 가설 H_0:β_j=0임을 검정해보자.

(1) 검정통계량

귀무가설 하에서 다음과 같은 검정통계량 t는 자유도가 n-p-1인 t-분포를 따른다.

$$t=\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\sim t(n-p-1)$$

2) 조정된 결정계수

모형에 대한 적합성을 확인하기 위하여 단순회귀분석과 마찬가지로 결정계수 R^2를 이용할 수 있다. 그러나 결정계수는 독립변수의 수가 늘어나면 증가하는 경향이 있다. 즉, 종속변수의 변동을 설명하는데 영향을 주지 못해도 독립변수를 추가하면 결정계수가 증가하는 것이다. 이를 보정한 방법이 조정된 결정계수(R^2_{adj})로 다음과 같이 정의된다.

$$R^2_{adj}=1-\left[(1-R^2)\frac{n-1}{n-p-1}\right]$$

3) 가변수 설정

회귀분석에서는 독립변수가 연속형(continuous variable)인 경우가 많지만, 때때로 범주형(categorical variable) 변수를 포함해야 하는 경우가 있다. 예를 들어, 성별(남/여), 지역(서울/부산/대구), 직업 유형(사무직/생산직/서비스업)과 같은 변수는 연속적인 수치로 측정되지 않는다. 회귀분석에서는 이러한 범주형 변수를 수치형 변수로 변환하여 모델에 포함해야 하는데, 이때 사용하는 방법이 가변수(dummy variable) 설정이다. 가변수는 0과 1의 값을 갖는 이진 변수(binary variable)로, 특정 범주에 속하는지를 나타낸다. 범주형 변수의 범주가 k개인 경우 가변수는 k-1개의 이진 변수를 생성한다.

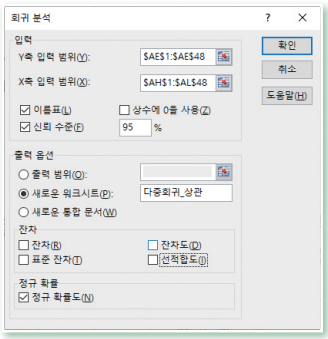
이때 k개의 범주 중 하나를 기준범주(reference category)로 결정한 후 나머지 k-1개의 범주에 대해 이진 변수를 만들면, 회귀계수는 기준범주 대비 해당 범주의 비교로 해석이 가능하다.

[실습] 다중회귀분석(ex_survey_data.xlsx)

이번 실습에서는 Excel을 통해 다중회귀분석을 진행하여 각 변수들이 혈중 크레아티닌 농도에 미치는 영향을 평가하고, 어떤 요인이 가장 중요한지 확인해보자.

① 데이터 → 데이터 분석 → 회귀분석을 클릭한다.

② [입력 Y 범위]란에는 혈중 크레아티닌 농도 데이터를 선택하고 [입력 X 범위]란에는 T_WAIST, T_PULSE, T_HBA1C, T_ALT, T_LDL 데이터를 선택한다. 단순회귀분석과 마찬가지로 [이름표], [신뢰수준], [정규확률도]를 선택한다.



③ 본 실습에서는 새로운 워크시트에 출력 결과를 출력해보자. 시트명은 "다중회귀_Creatinine"으로 한다. 마지막으로 확인을 눌러 분석을 진행한다.

④ 주요 출력 결과는 아래와 같다. Regression Statistics에서 Multiple R은 독립변수와 종속변수 간의 상관관계를 나타내는 상관계수로, 이번 분석에서는 0.6357로 나타났다. 이는 독립변수와 종속변수 간에 약한 양의 상관관계가 존재함을 의미한다. R^2은 결정 계수로, 독립변수들이 종속변수의 변동성을 얼마나 설명하는지 나타내며, 이번 실습에서는 0.4041로 확인되었다. 이는 독립변수들이 종속변수의 변동성을 40.41%를 설명한다는 뜻이다. Adjusted R^2은 수정된 결정 계수로, 모델의 변수 개수를 고려하여 R^2 값을 조정한 것이다. 값이 낮은 것은 추가 변수 중 일부가 기여도가 낮다는 것을 의미한다. ANOVA 표에서 유의한 F는 F-검정 결과를 의미한다. p-value가 0.05보다 작으므로 유의수준 0.05에서 전체 모델은 유의미하다고 해석할 수 있다. 아래 회귀계수 추정치와 검정 결과를 보면, T_WAIST가 종속변수에 유의미한 영향을 미치는 변수로 나타났다. 이는 허리둘레가 혈액 Creatinine 농도에 영향을 미친다는 것을 의미한다. 반면, 나머지 변수들

은 유의미하지 않은 것으로 나타났으며, 이는 해당 변수들이 모델에서 종속변수의 변동성을 설명하는 데 기여하지 않는다는 것을 시사한다.

	A	B	C	D	E	F	G	H	I
1	요약 출력								
2									
3	회귀분석 통계량								
4	다중 상관계수	0.635694385							
5	결정계수	0.404107351							
6	조정된 결정계수	0.331437515							
7	표준 오차	0.138797264							
8	관측수	47							
9									
10	분산 분석								
11		자유도	제곱합	제곱 평균	F 비	유의한 F			
12	회귀	5	0.535642	0.107128	5.560868	0.000533			
13	잔차	41	0.789852	0.019265					
14	계	46	1.325494						
15									
16		계수	표준 오차	t 통계량	P-값	하위 95%	상위 95%	하위 95.0%	상위 95.0%
17	Y 절편	-0.477919017	0.441306	-1.08297	0.285152	-1.36915	0.413316	-1.36915	0.413316
18	T_WAIST	0.008456057	0.002128	3.974431	0.000279	0.004159	0.012753	0.004159	0.012753
19	T_PULSE	-0.002467484	0.002581	-0.9559	0.344727	-0.00768	0.002746	-0.00768	0.002746
20	T_HBA1C	0.127711371	0.073136	1.746213	0.088263	-0.01999	0.275413	-0.01999	0.275413
21	T_ALT	-9.18528E-05	0.000402	-0.22843	0.820449	-0.0009	0.00072	-0.0009	0.00072
22	T_LDL	0.000901665	0.000705	1.278987	0.208093	-0.00052	0.002325	-0.00052	0.002325

4. 회귀진단

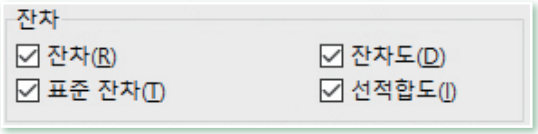
회귀분석에서 모형이 적절한지 평가하는 과정이 필요하다. 왜냐하면 적절하지 못한 모형에서 추정된 회귀계수에 대한 추론은 적절하다고 보기 어렵기 때문이다. 회귀진단(regression diagnostics)은 회귀모형이 적절하게 설정되었는지 확인하고, 가정이 충족되는지 평가하는 과정이다. 회귀진단을 통해 모형의 신뢰성을 높이고, 개선이 필요한 부분을 식별할 수 있다. 회귀진단은 주로 잔차분석(residual analysis)를 통해서 평가된다.

1) 잔차 분석

잔차 분석은 회귀모형의 가정을 평가하기 위해 수행된다. 주요 가정으로는 오차의 정규성, 오차의 등분산성, 오차의 독립성이 있다. 이를 평가하기 위해 개별 관측치에서 예측값을 뺀 잔차를 이용한다. x축을 예측값으로 하고 y축은 잔차나 표준화잔차로 하여 산점도로 시각화한 것을 잔차그림(residual plot)이라고 한다. 잔차그림을 통해 오차의 정규성이나, 분산의 동질성, 독립성 등을 확인한다.

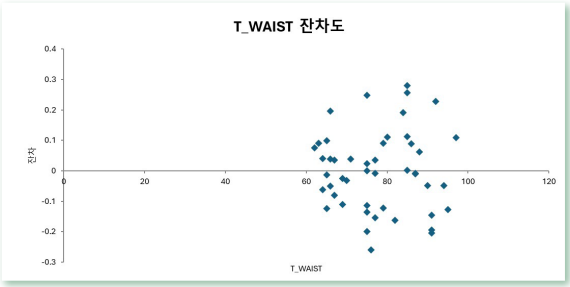
[실습] 잔차 분석

Excel을 통해 잔차 분석을 하기 위해서는 회귀분석의 대화 상자에서 잔차 옵션을 모두 선택하면 된다.



아래의 그림은 다중회귀분석 결과로부터 얻은 잔차 그림이다. 잔차가 모두 -0.2와 0.2사이에 골고루

퍼져있는 것을 확인할 수 있다. 잔차가 특정 패턴 없이 무작위로 분포된 것으로 보아 등분산성 가정을 만족한다고 볼 수 있다. 따라서, T_WAIST 변수는 잔차 가정을 만족한다고 볼 수 있다.



2) 다중공선성 확인

다중회귀분석에서 다중공선성(multicollinearity)는 중요한 문제 중 하나이다. 다중공선성은 독립변수들 간에 높은 상관관계가 있을 때 발생하며, 이는 회귀계수의 신뢰도를 떨어뜨리고 모형의 해석력을 약화시킬 수 있다. 다중공선성을 확인하기 위해 분산확대인자(variance inflation factor, *VIF*) 값을 사용한다. 독립변수 x_j에 대해 분산확대인자는 VIF_j이며, VIF_j가 10이상일 때 x_j와 다른 독립변수 간에 높은 상관관계가 있다고 볼 수 있다. 이 경우 다중공선성을 크게 하는 독립변수를 모형에서 제거하거나 변수 변환을 하는 등의 조치를 취해야 한다.

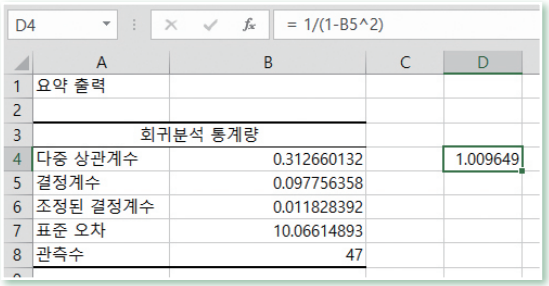
[실습] 다중공선성 확인

① 먼저 설명변수 x_1에서의 분산확대인자 VIF를 구한다. 이는 x_1을 반응변수로, 나머지 설명변수은 그대로 설명변수로 넣어 회귀분석을 해야 한다.

요약 출력	
회귀분석 통계량	
다중 상관계수	0.312660132
결정계수	0.097756358
조정된 결정계수	0.011828392
표준 오차	10.06614893
관측수	47

② 결정계수는 0.0978로 계산되었으며, 분산확대인자 *VIF*는 공식에 따라 다음과 같이 구할 수 있다. 빈 셀을 클릭한 후 “=1/(1-B5^2)” 수식을 입력하면 *VIF* 값이 계산되며, x_1에 대한 *VIF* 값은 1.01으로 나타나는 것을 확인할 수 있다. 이와 동일한 방식으로 다른 설명변수에 대해서도 분산확대인자 VIF_j를 구할 수 있으며, 일반적으로 VIF_j가 10 이상일 때, x_j와 다른 설명변수 간에 다중공선

성이 존재한다고 판단한다. 따라서, T_WAIST 변수는 다중공선성이 없는 것으로 판단할 수 있다.



D4 = 1/(1-B5^2)

	A	B	C	D
1	요약 출력			
2				
3	회귀분석 통계량			
4	다중 상관계수	0.312660132		1.009649
5	결정계수	0.097756358		
6	조정된 결정계수	0.011828392		
7	표준 오차	10.06614893		
8	관측수	47		

제2절 보건통계학의 응용

1 비모수 검정

모수적 방법(parametric method)은 표본이 추출된 모집단이 정규분포나 이항분포 등 특정한 분포를 따른다고 가정하고 분석을 수행하는 통계적 방법이다. 예를 들어, 앞장에서 다룬 t-검정과 분산분석은 모집단이 정규분포를 따른다는 가정하에 적용되며, 이러한 분석은 가정이 충족될 때만 적절하다.

반면, 비모수적 방법(nonparametric method)은 모집단이 특정 분포를 따른다는 가정을 하지 않는다. 대신, 모집단 분포가 연속적이거나 대칭적이라는 최소한의 가정만을 바탕으로 한다. 비모수 검정은 자료의 분포가 모수적 방법의 가정을 만족하지 않거나, 표본의 크기가 작아 중심극한정리를 적용하기 어려운 경우 유용한 대안이 된다.

예를 들어, "성별에 따라 수축기 혈압이 차이가 있는가?"라는 질문에 답하기 위해 표본을 추출했다고 가정하자. 이 경우, 적절한 통계분석을 수행하기 위해서는 먼저 자료의 분포를 확인해야 한다. 수축기 혈압은 연속형 데이터이므로, 이를 두 집단 간 연속형 데이터 비교 문제로 정의할 수 있다. 연속형 데이터의 집단 간 비교를 위해서는 자료가 정규분포를 따르는지 확인한 후,

- 정규성을 만족하면 모수적 방법을 적용하고,
- 정규성을 만족하지 않으면 비모수적 방법을 고려해야 한다.

비모수적 방법은 모수적 방법에 비해 분포에 관한 가정이 완화되거나 거의 없다. 특히 소표본인 경우에도 안정적인 특징이 있다. 비모수적 방법들은 대부분 부호(sign) 또는 상대크기의 순위(rank) 등 직관적으로 이해하기 쉬운 정보를 사용한다. 또한 많은 경우에 모수적 방법과 비교하여 표본의 크기가 큰 경우에 검정력과 효율이 동일한 것으로 알려져 있다. 그러나 비모수적 방법은 이론 전개가 복잡하고 어려우며, 특정한 모집단 하에서는 모수적 방법에 비해 효율이 떨어진다는 특징이 있다.

비모수 검정에서는 원자료의 수치값 자체보다 자료의 부호 또는 순위 정보를 이용하여 검정통계량

을 구성하므로, 이상치(outlier)에 덜 민감하다. 또한 순위는 순서형 자료로 항목 간 순서만 있고 수치 간 간격은 의미 없기 때문에 평균이나 표준편차와 같은 연산은 부적절하다. 따라서, 비모수 검정에서는 평균 비교 대신 위치 모수(location parameter)비교 또는 중위수(median) 비교라고 한다.

이 장에서는 독립적인 두 집단에서 연속형 자료를 비교하는 비모수 검정 방법인 윌콕슨 순위합검정(wilcoxon rank-sum test)을 다룬다. 또한, 짝지어진 표본에서 사용할 수 있는 비모수 검정 방법으로 부호 검정(sign test)과 윌콕슨 부호순위 검정(wilcoxon signed-rank test)을 설명한다. 마지막으로, 세 개 이상의 집단을 비교할 때 활용할 수 있는 비모수 검정 방법인 크루스칼-왈리스 검정(kruskal-wallis test)에 대해 다룬다. 비모수 검정은 Excel에서 분석 기능을 제공하지 않는다. 따라서, R, SPSS, SAS와 같은 통계 프로그램을 이용하여 분석을 수행해야 한다.

1. 두 독립 집단 간 연속형 자료의 비교: 윌콕슨 순위합검정

윌콕슨 순위합검정(wilcoxon rank-sum test)은 두 개의 독립된 집단 간 위치모수(location parameter)의 비교 검정을 하기 위한 비모수적 방법이다. 이 검정은 연속형 데이터를 대상으로 하며, 맨-휘트니 검정(Mann-Whitney U Test)과 결과적으로 동일한 검정 방법이다. 윌콕슨 순위합검정은 특정 분포를 가정하지 않으며, 관찰값의 실제 크기보다 순위(rank) 정보에 근거하여 검정을 한다.

두 집단 각각 n_1개와 n_2개의 표본이 할당되었다고 하자. $X_1, \cdots, X_{n_1}$을 위치모수가 θ_1인 분포 $F(x)$로부터의 확률표본이라고 하고, $Y_1, \cdots, Y_{n_2}$을 위치모수가 θ_2인 분포 $G(y)$로부터의 확률표본이라고 할 때, 윌콕슨 순위합검정의 가설은 다음과 같다(양측검정).

H_0: $\theta_1=\theta_2$ 두 집단간 중위수(또는 위치모수)가 차이가 없다.
H_1: $\theta_1\neq\theta_2$ 두 집단간 중위수(또는 위치모수)가 차이가 있다.

먼저 두 집단의 데이터를 합쳐 총 $n_1+n_2=n$개의 데이터에 대해 오름차순으로 나열한다. 작은 값부터 차례로 순위를 매긴다. 만약 동일한 관찰값들(동점)이 있다면 해당 값들은 해당 값들의 순위의 평균값을 부여한다. 그리고 첫 번째 집단의 순위의 합 R_1을 계산하고 두 번째 집단의 순위의 합 R_2를 계산한다. n개 자료의 순위 합은 항상 $n(n+1)/2$가 된다. 따라서 $R_1+R_2=n(n+1)/2$가 성립한다.

확률표본 $X_1, \cdots, Xn_1$의 표본크기 n_1과 확률표본 $Y_1, \cdots, Yn_2$의 표본크기 n_2이 충분히 큰 경우(대표본) 정규근사를 이용할 수 있다. 이때 검정통계량 R_1은 근사적으로 정규분포를 따른다. 귀무가설하에 R_1의 평균과 분산은 다음과 같이 주어진다.

$$E(R_1)=\frac{n_1(n_1+n_2+1)}{2},\ Var(R_1)=\frac{n_1n_2(n_1+n_2+1)}{12}$$

이를 이용하여 다음과 같이 두 집단 간 중위수의 차이에 대한 검정을 할 수 있다.

① 두 집단의 자료를 합친 후(혼합표본 $n_1+n_2=n$) 오름차순으로 나열한다. 각 관찰값에 순위를 부여한다.
② 만약 값이 동일한 관찰값들이 있다면 (signed-rank test와 같이) 순위의 평균값으로 할당한다.
③ 집단 1과 집단 2의 순위합 R_1, R_2를 계산한다.
④ 검정통계량(대표본인 경우)

$R_1 \neq n_1(n_1+n_2+1)/2$이면서 동점 자료가 없는 경우,

$$Z=\left[\left|R_1-\frac{n_1(n_1+n_2+1)}{2}\right|-\frac{1}{2}\right]/\sqrt{(\frac{n_1n_2}{12})(n_1+n_2+1)}$$

$R_1 \neq n_1(n_1+n_2+1)/2$이면서 동점 자료가 있는 경우,

$$Z=\left[\left|R_1-\frac{n_1(n_1+n_2+1)}{2}\right|-\frac{1}{2}\right]/\sqrt{(\frac{n_1n_2}{12})(n_1+n_2+1-\sum_{i=1}^{g}\frac{t_i(t_i^2-1)}{(n_1+n_2)(n_1+n_2-1)})}$$

$R_1 \neq n_1(n_1+n_2+1)/2$인 경우, $Z=0$
여기서 t_i는 g개의 동점자료 그룹이 있을 때, i번째 동점자료 그룹에서 자료의 수이다.

⑤ 가설검정 결과

a. 양측검정
- H_0: $\theta_1=\theta_2$ vs $H_1=\theta_1 \neq \theta_2$일 때 $|Z|>z_{1-\alpha/2}$이면 귀무가설 기각

b. 단측검정
- H_0: $\theta_1=\theta_2$ vs $H_1=\theta_1<\theta_2$일 때 $Z<-z_{1-\alpha}$이이면 귀무가설 기각
- H_0: $\theta_1=\theta_2$ vs $H_1=\theta_1>\theta_2$일 때 $Z>z_{1-\alpha}$이면 귀구가설 기각

여기서 $Z_{1-\alpha}$는 표준정규분포에서 하위 1-α 확률에 해당하는 임곗값이다.

[실습] 데이터 불러오기

Excel에서는 비모수 검정 기능을 제공하지 않는다. 따라서 비모수 검정을 수행하려면 R 프로그램을 활용해야 한다. 먼저, 예제 데이터 "ex_survey_data.xlsx" 파일을 R에서 불러와 저장해야 한다. 이를 위해 readxl 패키지를 사용하면 xlsx 확장자 파일을 불러와 "wrs"라는 이름의 데이터로 저장할 수 있다.

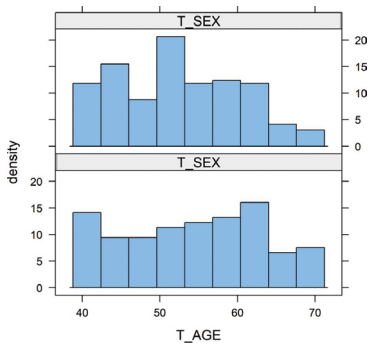
R code	> # 데이터 불러오기 > library (readxl) > wrs <- read_excel("ex_survey_data.xlsx")

[실습] 정규성 확인

연령의 분포가 성별에 따라 차이가 있는지 검정해보자. 이를 위해 먼저 연령이 정규분포를 따르는지 확인해야 한다. 그래프를 활용하여 분포를 시각적으로 확인한 후, Shapiro-Wilk 검정을 수행하여 정규성을 검정한다.

다음 R 코드는 lattice 패키지를 사용하여 wrs 데이터에서 T_AGE (나이) 변수를 성별(T_SEX)에 따라 나누어 히스토그램을 그리는 코드이다. 또한, dplyr 패키지를 활용하여 wrs 데이터프레임을 T_SEX (성별)별로 그룹화한 후, Shapiro-Wilk 정규성 검정을 수행하여 결과를 정리하는 코드도 포함되어 있다.

R code	> # 히스토그램 > library(lattice) > histogram(~T_AGE|T_SEX, data = wrs, ylab = "density", layout = c(1,2)) > # 정규성 검정 > library(dplyr) > shapiro_result <- wrs %>% + group_by(T_SEX) %>% + summarise(+ T_AGE.statistic = unlist(shapiro.test(T_AGE)$statistic), + T_AGE.p.value = unlist(shapiro.test(T_AGE)$p.value) +)

결과	 > print(shapiro_result) # A tibble: 2 x 3 T_SEX T_AGE.statistic T_AGE.p.value ⟨db1⟩ ⟨db1⟩ ⟨db ⟩ 1 1 0.956 0.00137 2 2 0.969 0.000247

히스토그램을 확인한 결과, 연령은 정규분포의 형태를 따르지 않는 것으로 보인다. 또한, Shapiro-Wilk 검정 결과, 두 집단 모두에서 유의수준 5%에서 "H_0:정규분포를 따른다"는 귀무가설을 기각하였다. 따라서 연령이 정규분포를 따른다고 보기 어렵다. 이에 따라, 성별에 따른 연령의 차이를 검정할 때 비모수 검정을 수행하는 것이 적절하다.

[실습] 윌콕슨 순위합검정

R code	> # 윌콕슨 순위합검정 > wilcox.test(T_AGE ~ T_SEX, data = wrs)
결과	Wilcoxon rank sum test with continuity correction data: T_AGE by T_SEX W 11510, p-value 0.08715 alternative hypothesis: true location shift is not equal to 0

p-value=0.087로 유의수준 5%에서 "H_0: 남녀 간 연령의 중위수는 같다"라는 귀무가설을 기각할 수 없다. 따라서 남녀 간 연령의 중위수는 같다고 할 수 있다.

2. 짝지은 표본에서 연속형 자료의 비교: 윌콕슨 부호순위 검정

짝지은 표본에 적합한 모수적 방법으로는 대응표본 t-검정(paired t-test)이 있다. 그러나 모집단의 정

규성 가정을 충족하기 어려운 경우에는 비모수적 방법을 사용할 수 있다. 짝지은 표본의 비모수적 방법으로는 부호 검정(sign test)과 윌콕슨 부호순위 검정(Wilcoxon signed rank test)이 있다.

1) 부호 검정

부호 검정은 짝지은 표본 A집단과 B집단이 있을 때, 대응된 개체에서 A집단의 값이 B집단의 값보다 크거나, 작거나, 같은지는 알 수 있지만 상대적인 점수의 크기는 알 수 없는 경우에 가설검정이다.

(1) 가설

H_0:Δ=0 vs H_1:Δ≠0

여기서 Δ는 두 대응표본의 차이에 대한 모집단 중위수를 의미한다.

(2) 검정통계량

검정통계량 c는 대응표본의 차이값 d_i 중에서 양수의 개수이며, n>20인 경우 통계량 C는 귀무가설 하에서 평균이 $\frac{n}{2}$, 분산이 $\frac{n}{4}$인 정규분포에 근사하는 성질을 이용하여 가설검정을 수행한다. n은 d_i가 0이 아닌 d_i의 개수이다.

$$C=d_i>0\text{인 } d_i\text{의 수} \sim N(\frac{n}{2}, \frac{n}{4}), (n>20)$$

(3) 가설검정 결과

$$C>c_2=\frac{n}{2}+\frac{1}{2}+z_{1-\alpha/2}\sqrt{\frac{n}{4}} \text{ or } C<c_1=\frac{n}{2}+\frac{1}{2}+z_{1-\alpha/2}\sqrt{\frac{n}{4}}\text{이면, } H_0 \text{ 기각}$$

2) 윌콕슨 부호순위 검정

부호 검정은 차이값의 부호 정보만을 이용한다. 그러나 차이에 대해 정성적인 평가가 가능하고, 이를 수치값으로 나타낼 수 있다면, 차이값의 정보를 활용하는 것이 바람직하다. 이 경우에는 윌콕슨 부호순위 검정을 사용하여 다음과 같은 가설을 검정할 수 있다.

(1) 가설

$$H_0: \Delta=0 \text{ vs } H_1: \Delta\neq 0$$

여기서 Δ는 두 짝지은 표본의 차이에 대한 모집단 중위수를 의미한다.

(2) 검정 절차

① 차이 d_i의 절대값을 순서대로 정렬한다.

② $d_i=0$인 관찰값은 무시하고 가장 작은 절대값에 1을 부여하며 가장 큰 절대값 순서대로 순위를 부여한다.

③ 만약 절대값이 같은 관찰값들에 대해서는 평균순위를 부여한다.

④ 양수인 차이값의 순위를 합하여 T_+이라 하고 음수인 차이값의 순위를 합하여 T_-라 한다.

T_+=부호가 양수인 차이값 d_i의 순위들의 합

T_-=부호가 음수인 차이값 d_i의 순위들의 합

⑤ 검정통계량 T는 T_+와 T_- 중 작은 값으로 정의하며, 검정통계량 T의 분포로부터 p-value를 계산한다. 검정통계량 T의 분포는 복잡하므로 이 책에서는 다루지 않는다. 통계 소프트웨어를 활용하여 윌콕슨 순위합검정을 수행하면 p-value를 제공해 주므로 이를 이용한다.

3. 여러 집단의 비교: 크루스칼-왈리스 검정

앞에서 비교하는 집단이 셋 이상인 경우 연속형 자료에 대해 평균을 비교하는 모수적 방법인 분산분석을 소개하였다. 분산분석의 가정은 연속형 자료가 정규분포를 따른다는 가정이 있다. 만약 연속형 자료의 분포가 정규분포를 따른다고 보기 어려운 경우 완화된 가정으로 비모수적 방법을 사용할 수 있다. 비교하는 집단이 셋 이상인 경우에 연속형 자료의 중위수를 비교하는 비모수적 방법은 크루스칼-왈리스 검정(Kruskal-Wallis test)이다. 크루스칼-왈리스 검정은 앞 절에서 다룬 윌콕슨의 순위합검정의 확장으로서, 크루스칼-왈리스 검정에서 비교하는 집단의 수가 2일 때는 윌콕슨의 순위합검정과 동일하다.

1) 가설

H_0: 모집단에서 집단들의 분포는 서로 같다.

H_1: 모집단에서 집단들의 분포는 서로 같지 않다.

2) 검정통계량

표본크기가 n_1, n_2, …, n_k인 집단을 비교한다고 가정하자. 모든 $n(=n_1+n_2+\cdots n_k)$개의 관측값을 순서대로 나열하여 순위를 매긴 후 각 집단별로 순위의 합을 계산한다. 이 순위합을 R_1, R_2, …, R_k라고 하자. 이때 검정통계량은 다음과 같으며, 자유도가 k-1인 카이제곱 분포를 따른다. 만약 동일한 순위를 갖는 값이 있는 경우 수정된 검정통계량을 사용한다.

$$H=\frac{12}{n(n+1)}\Sigma\frac{R_i^2}{n_i}-3(n+1)\sim X_\alpha^2(k-1)$$

[실습] 세 집단의 중위수 비교

연령의 분포가 흡연 여부에 따라 차이가 있는지 검정해보자. "ex_survey_data.xlsx" 파일 내에서 흡연 여부(T_SMOKE) 변수는 다음과 같이 분류되어 있다. 1=아니오, 2=예(과거흡연), 3=예(현재흡연). 세 집단 간 연령의 차이가 있는지 확인하기 위해, 앞서 진행한 방법과 동일하게 정규성을 검정한 후, 적절한 검정을 수행해야 한다. 정규성 검정을 거친 후, 다음 R 코드를 실행하면 크루스칼-왈리스 검정을 수행할 수 있다.

R code	> # 크루스칼-왈리스 검정 > kruskal.test(T_AGE~T_SMOKE, data = wrs)
결과	Kruskal-wallis rank sum test data: T_AGE by T_SMOKE Kruskal-Wallis chi-squared = 6.5773, df = 2 p-value = 0.0373

p-value = 0.0373으로 유의수준 5%에서 "H_0: 흡연 여부에 따라 연령의 중위수가 같다"를 기각하였다. 따라서 세 집단(비흡연, 과거흡연, 현재흡연) 간 연령의 중위수는 다르다고 할 수 있다.

2 로지스틱 회귀분석

로지스틱 회귀분석(logistic regression)은 이분형(binary) 종속변수(예: 질병 유무, 성공/실패, 합격/불합격 등)에 대한 회귀모형을 적합할 때 사용된다. 일반적인 선형회귀모형은 종속변수가 연속형인 경

우 여러 독립변수들과 종속변수 간의 관계를 모형화한다. 하지만, 종속변수가 0과 1로 구분되는 범주형 변수를 가질 때는 선형회귀를 그대로 적용할 수 없다. 선형회귀를 그대로 사용할 경우, 예측값이 0과 1 사이를 벗어날 수 있기 때문이다.

이를 해결하기 위해, 로지스틱 회귀분석에서는 종속변수의 확률을 예측하는 방식을 사용한다. 즉, 어떤 사건이 발생할 확률을 예측하는 모델이다. 종속변수 y의 결과값을 0과 1로 할 때, 종속변수 y의 평균은 다음과 같다.

$$E(Y)=1\times p(Y=1)+0\times p(Y=0)=p(Y=1)$$

여기서 종속변수 y에 영향을 주는 독립변수들이 있는 경우 반응변수의 성공확률은 이러한 독립변수들의 선형결합으로 다음과 같이 선형모형을 설정할 수 있다.

$$p_x=\alpha+\beta_x$$

그러나 p_x는 그림 2-2-1과 같이 독립변수 x와 비선형관계이므로, 위와 같이 선형모형을 설정하게 되는 경우 p_x의 범위는 0과 1사이를 넘어가게 된다. 이러한 어려움을 피하기 위해 p_x 대신 로짓(logit) 변환된 값을 종속변수로 사용하여 다음과 같은 모형식을 설정할 수 있다. 이를 로지스틱 회귀모형(logistic regression model)이라고 한다.

$$\ln\frac{p_x}{1-p_x}=\alpha+\beta_x$$

여기서 $p_x=\frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)}$, $1-p_x=\frac{1}{1+\exp(\alpha+\beta x)}$와 상등한다.

로지스틱 회귀모형에서 회귀계수 α, β에 대하여 선형회귀분석에서 사용하는 최소제곱법(Ordinary Least Squares, OLS)으로 추정할 수 없다. 이는 로지스틱 회귀모형이 종속변수의 범주형 특성을 반영하여 확률을 예측하는 방식이기 때문이다. 따라서, 회귀계수를 추정하기 위해 우도 함수(likelihood function)를 최대화하는 최대우도추정(Maximum Likelihood Estimation, MLE)을 사용해야 한다. 로지스틱 회귀모형에서 최대우도추정량을 구하기 위해서는 뉴턴-랩슨(Newton-Raphson) 방법과 같은 반복 최적화 알고리즘을 적용해야 한다. 이러한 계산 과정은 반복적인 수치 최적화를 필요로 하므

로, Excel에서는 로지스틱 회귀분석 기능을 기본적으로 제공하지 않는다. 대신, R, Python, SPSS, SAS 등의 통계 소프트웨어를 활용하여 로지스틱 회귀계수를 추정할 수 있다.

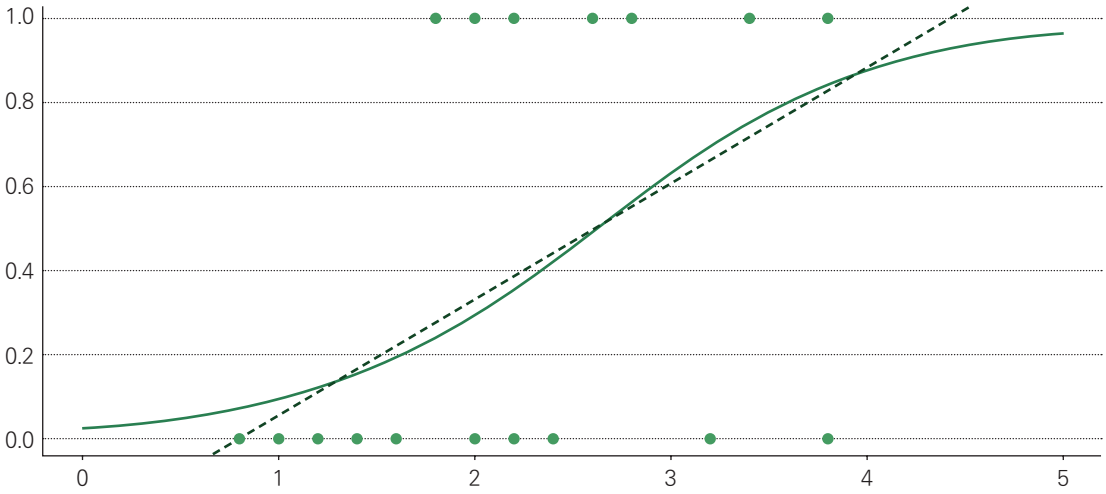


그림 1-2-1. 연속형 독립변수 x와 y=1일 확률 곡선

이 교재에서는 로지스틱 회귀분석의 개념과 결과 해석을 중심으로 다룬다. 회귀계수의 추정 방법과 수학적 유도 과정에 대한 상세한 설명보다는, 로지스틱 회귀모형이 어떤 원리로 동작하는지 이해하고, 분석 결과를 올바르게 해석하는 방법에 초점을 맞춘다. 이를 통해, 로지스틱 회귀분석을 실무에서 활용할 수 있도록 하는 데 목적을 둔다.

1. 단순 로지스틱 회귀분석

단순 로지스틱 회귀분석(simple logistic regression)은 하나의 독립변수 X가 종속변수 Y에 미치는 영향을 분석하는 회귀모형이다. 이때 종속변수 Y는 이항(binary) 변수로, 0 또는 1의 값을 갖는다. 역학 연구에서 활용되는 예시로는, 단순 로지스틱 회귀분석은 특정 위험 요인(독립변수)이 질병 발생(종속변수)에 미치는 영향 분석이 있다. 로지스틱 회귀분석을 통해 역학연구에서는 질병 발생 가능성을 예측하고, 특정 위험 요인의 영향을 정량적으로 평가할 수 있다.

1) 단순 로지스틱 회귀모형

$$\ln\frac{p_i}{1-p_i}=\beta_0+\beta_1 x_i$$

여기서 p_i는 i번째 개체의 Y=1일 확률이며, β_0, β_1은 회귀계수이다. 단순로지스틱 회귀모형 정리하면 다음과 같다.

$$p_i=\frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}$$

즉, 독립변수가 변화함에 따라 Y=1일 확률이 어떻게 변하는지를 나타내는 함수이다. 여기서 β_1이 양수이면 독립변수가 커질수록 Y=1일 확률도 증가하며, β_1이 음수이면 독립변수가 커질수록 Y=1일 확률은 감소한다.

2) 회귀계수의 해석

(1) 독립변수가 연속형인 경우

독립변수가 x일 때 Y=1에 대한 확률을 Px라 하고, 독립변수가 x+1일 때 Y=1에 대한 확률을 Px_{+1}이라고 하자. 이때 독립변수 x가 한단위 증가할 때(x+1), 오즈비는 $\exp(\beta_1)$과 같다.

$$\frac{\frac{P_{x+1}}{1-P_{x+1}}}{\frac{P_x}{1-P_x}}=\frac{\exp(\beta_0+\beta_1(x+1))}{\exp(\beta_0+\beta_1 x)}=\exp(\beta_1)$$

따라서 회귀계수에 지수함수를 취하여 오즈비로 해석이 가능하다. 예를 들어, $\widehat{\beta_1}$=0.8일 때 독립변수가 한 단위 증가할 때 Y=1일 오즈는 $e^{0.8}$으로 약 2.23이다. 따라서 독립변수가 한 단위 증가할 때 Y=1일 오즈는 2.23배가 된다고 해석할 수 있다.

(2) 독립변수가 범주형인 경우

세 가지 처리 A, B, C 방법에 대하여 이분형 종속변수 Y에 대해 로지스틱 회귀모형은 다음과 같이

가변수 D_{11}, D_{12}를 이용하여 표현할 수 있다.

$$\ln\frac{P}{1-P}=\beta_0+\beta_1\times D_{11}+\beta_2\times D_{12}$$

처리방법은 범주가 A, B, C 세 개이므로 두 개의 가변수를 모형에 포함한다. 이때 기준범주를 C라고 할 때 가변수 D_{11}은 처리 A를 받았으면 1, 그렇지 않으면 0으로 정의한다. D_{12}는 처리 B를 받았으면 1, 그렇지 않으면 0으로 정의한다. 만약 C를 받았다면, D_{11}=0, D_{12}=0이 된다. 여기서 기준범주 처리 C 대비 처리 A의 Y=1에 대한 오즈비를 구하면 다음과 같이 $\exp(\beta_1)$과 같다.

$$\frac{P_A}{1-P_A}/\frac{P_C}{1-P_C}=\frac{\exp(\beta_0+\beta_1)}{\exp(\beta_0)}=\exp(\beta_1)$$

마찬가지로 기준범주 처리 C 대비 처리 B의 Y=1에 대한 오즈비를 구하면 $\exp(\beta_2)$와 같다.

$$\frac{P_B}{1-P_B}/\frac{P_C}{1-P_C}=\frac{\exp(\beta_0+\beta_2)}{\exp(\beta_0)}=\exp(\beta_2)$$

즉, 처리 A를 받았을 때 Y=1일 오즈는 처리 C를 받았을 때의 오즈의 $\exp(\beta_1)$배라고 해석할 수 있다. $\exp(\beta_2)$도 동일한 방법으로 해석한다.

2. 다중 로지스틱 회귀분석

이제 p개의 독립변수 $x_1, x_2, \cdots, x_p$를 고려하자. 종속변수가 0 또는 1의 값을 갖는 이분형일 때 다중 로지스틱 회귀모형(multiple logistic regression model)은 다음과 같다. 여기서 p_i는 i번째 개체의 Y=1에 대한 확률이다.

$$\ln\frac{p_i}{1-p_i}=\beta_0+\beta_1x_{i1}+\cdots+\beta_px_{ip}$$

여기서 p_i는 Y=1일 확률이며, β_0, β_1, … , β_p는 회귀계수이다. 다중 로지스틱 회귀분석에서 회귀계수의 해석은 단순 로지스틱 회귀분석에서의 회귀계수와 같이 오즈비를 계산하여 해석하는 것이 동일하지만 다소 차이가 있다. 단순 로지스틱 회귀분석에서 회귀계수는 한 독립변수가 종속변수에 미치는 영향으로 직접 해석할 수 있지만, 다중 로지스틱 회귀분석에서는 다른 독립변수들의 영향을 통제한 상태에서 특정 독립변수가 종속변수에 미치는 순수한 효과를 나타낸다는 점에서 차이가 있다. 즉, 단순 로지스틱 회귀분석에서 회귀계수 β_1에 대한 해석은 오즈비(OR)로 해석되며, 독립변수가 한 단위 증가할 때 Y=1에 대한 오즈가 $\exp(\beta_1)$배 증가하거나 감소함을 의미한다. 그러나 다중 로지스틱 회귀분석에서는 회귀계수 β_j를 보정된 오즈비(adjusted OR)로 해석한다. 다른 독립변수들이 일정하게 유지된 상태에서 해당 독립변수가 종속변수에 미치는 순수한 효과를 나타낸다.

3. 회귀계수에 대한 가설검정

로지스틱 회귀분석에서 회귀계수에 대한 가설검정을 수행해보자. H_0:β_j=0이라는 귀무가설은 독립변수 x_j가 종속변수에 영향을 미치지 않는다는 것을 의미한다. 만약, 귀무가설을 기각하면 독립변수 x_j는 종속변수에 유의한 영향을 미친다는 것을 의미하는 것이다.

1) 가설

H_0:β_j=0 vs H_1:β_j≠0

2) 검정통계량

가설을 검정하기 위해 다음과 같은 왈드(Wald) 통계량을 사용한다. 왈드 통계량은 귀무가설 하에서 자유도가 1인 카이제곱 분포를 따르는 것을 이용하여 가설검정을 수행한다.

$$\left(\frac{\hat{\beta}_i}{SE_{\hat{\beta}_i}}\right)^2 \sim \chi^2_{\alpha}(1)$$

p-value가 0.05보다 작으면, 귀무가설 기각하고, 독립변수 x_j의 효과가 유의하다고 판단한다.

4. 모형의 적합도 평가

로지스틱 회귀모형이 데이터를 얼마나 잘 설명하는지를 평가하기 위해 다음과 같은 지표를 사용한다.

1) -2 로그 가능도(-2 log likelihood, -2LL)

모형이 데이터를 얼마나 잘 적합하는지를 나타내는 지표로, 값이 작을수록 더 좋은 모형을 의미한다.

2) Akaike 정보 기준(Akaike information criterion, AIC)

모형의 적합도를 평가하는 지표로, 값이 작을수록 더 좋은 모형으로 해석된다.

또한, 로지스틱 회귀모형의 예측값과 실제 관측값을 비교한 분류표(classification table)를 통해 모형의 예측 성능을 평가할 수 있다. 이를 통해 민감도(sensitivity), 특이도(specificity) 등의 예측 효율성 지표를 계산할 수 있으며, 이 값들이 높을수록 예측 성능이 우수하다고 해석할 수 있다.

[실습] 로지스틱 회귀분석

BMI가 고혈압에 영향을 미치는지 알아보자. "ex_survey_data.xlsx" 예제 데이터에는 고혈압 진단 여부(T_HTN)변수가 있으며, 1이면 "아니오", 2이면 "예" 이다. R에서 as.factor()을 이용하여 T_HTN 변수를 factor형으로 변환한다. glm( )함수를 이용하여 T_HTN (고혈압 여부)이 종속 변수이며, T_BMI (체질량지수, BMI)가 독립 변수인 단순 로지스틱 회귀모형에 적합해보자. 적합한 결과를 summary( ) 함수를 사용하여 요약한다.

R code	> # 고혈압 진단여부 > wrs$T_HTN <- as.factor(wrs$T_HTN) > fit <- glm(T_HTN ~ T_BMI , data = wrs, family = binomial("logit")) > summary(fit)

결과

```
〉 summary (fit)
Call:
glm(formula = T_HTN T_BMI, family binomial ("logit"), data = wrs)
Coefficients:
              Estimate Std. Error z value Pr (〉|z|)
(Intercept)   −3.7778   1.2687    −2.978   0.0029 **
T_BMI         0.1019    0.0519    1.963    0.0496 *
_ _ _
Signif. codes: 0 '***' 0.001 '**'  0.01 '*' 0.05 '·' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance:   308.37 on 299 degrees of freedom
Residual deviance:   304.45 on 298 degrees of freedom
AIC: 308.45

Number of Fisher Scoring iterations: 4
```

로지스틱 회귀분석 결과를 해석할 때는 기준이 되는 범주(reference)가 매우 중요한데, glm()함수는 자동으로 알파벳/숫자 순서에 따라 작은 값을 기준(reference)으로 설정한다. 즉, 여기서 T_HTN = 1이 기준(reference level)이 되고, T_HTN = 2를 예측하는 모델이 된다. summary(fit)을 이용하여 출력한 로지스틱 회귀분석 결과를 살펴보면, 왈드 검정(Wald test)결과 BMI의 p-value = 0.0496으로 유의수준 5%에서 "H_0:β=0"이라는 귀무가설을 기각한다. 따라서 BMI는 고혈압 진단여부에 유의한 영향을 미친다. 로지스틱 회귀모형의 적합식은 다음과 같다.

$$\ln\left(\frac{p_x}{1-p_x}\right)=0.0029+0.046BMI$$

이제 anova( )함수를 이용해 회귀계수의 유의성 검정을 수행해보자. anova( )함수에서 test = "LRT" 옵션을 사용하여 우도비 검정을 수행한다. GLM에서 특정 변수들이 모형에 포함될 필요가 있는지 검정할 때 유용하다.

R code	> #우도비검정 > anova(fit, test = "LRT") > #오즈비 > exp(fit$coefficients) > #오즈비의 신뢰구간 > exp(confint(fit))
결과	Analysis of Deviance Table Model: binomial, link: logit Response: T_HTN Terms added sequentially (first to last) Df Deviance Resid. Df Resid. Dev Pr(>Chi) NULL 299 308.37 T_BMI 1 3.9203 298 304.45 0.04771 * _ _ _ Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' . ' 0.1 ' ' 1 > exp(fit$coefficients) (Intercept) T_BMI 0.02287204 1.10727628 > exp(confint(fit)) Waiting for profiling to be done... 2.5% 97.5% (Intercept) 0.00179422 0.2635717 T_BMI 1.00102512 1.2277530

우도비 검정 결과 p-value=0.0477로 왈드 검정 결과와 유사하다. 따라서 BMI는 고혈압 진단여부에 영향을 미치는 요인임을 알 수 있다. 오즈비를 구해보면 exp(0.0496)=1.107(95% Cl: 1.001, 1.228)이다. 따라서 BMI가 한 단위 증가할수록 고혈압일 오즈는 1.107배이다. 즉, BMI가 증가할수록 고혈압일 가능성이 높다는 것을 시사한다.

이제 연령 변수를 추가로 고려해보자. 연령은 BMI에도 영향을 미치며, 고혈압에도 영향을 미칠 수 있는 변수이다. 따라서 BMI가 높은 사람들이 일반적으로 연령이 높아서 고혈압에 대한 오즈가 유의하게 나올 수 있으므로, 연령을 보정한 상태에서 BMI의 보정된 오즈비(adjusted OR)를 구해볼 필요가 있다. 이를 위해 연령을 추가하여 다중 로지스틱 회귀분석을 수행해보자.

아래는 R 코드와 그 결과이다.

R code	> #다중로지스틱 회귀분석 > fit <- glm(T_HTN ~ T_AGE + T_BMI , data = wrs, family = binomial("logit")) > summary(fit) > #오즈비 > exp(fit$coefficients) > #오즈비의 신뢰구간 > exp(confint(fit))
결과	> summary(fit) Call: glm(formula = T_HTN T_AGE + T_BMI, family binomial("logit"), data = wrs) Coefficients: Estimate Std. Error z value Pr()Izl) (Intercept) −7.70906 1.68394 −4.578 4.69e−06 *** T_AGE 0.08262 0.01916 4.313 1.61e−05 *** T_BMI 0.07686 0.05690 1.351 0.177 _ _ _ Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' · ' 0.1 ' ' 1 (Dispersion parameter for binomial family taken to be 1) Null deviance: 308.37 on 299 degrees of freedom Residual deviance: 283.88 on 297 degrees of freedom AIC: 289.88 Number of Fisher Scoring iterations: 4 > #오즈비 > exp(fit$coefficients) (Intercept) T_AGE T_BMI 0.000448741 1.086124982 1.079886058 > #오즈비의 신뢰구간 > exp(confint(fit)) Waiting for profiling to be done... 2.5% 97.5% (Intercept) 1.440832e−05 0.01082168 T_AGE 1.047117e+00 1.12907850 T_BMI 9.660100e−01 1.20837047

단순 로지스틱 회귀분석에서 유의하였던 BMI의 p-value = 0.177으로 유의수준 5%에서 유의하지 않으며, 보정된 오즈비는 1.08(95% CI: 0.966, 1.208)으로 95% 신뢰구간 또한 1을 포함하여 동일한 결과이다. 따라서 연령을 보정하였을 때 고혈압 진단여부에 BMI변수의 영향은 유의하지 않다.

3 생존분석

생존분석(survival analysis)은 시간(time)과 사건(event) 간의 관계를 분석하는 통계적 방법이다. 여기서 '생존'이라는 개념은 반드시 생물학적 생존을 의미하는 것이 아니라, 어떤 사건이 발생하기까지의 시간(time to event)을 분석하는 방법을 의미한다. 역학 연구에서 생존분석은 질병 발생의 위험, 재발 위험, 사망 위험과 관련된 연구 등에 활용된다. 예를 들어, 백신 접종 후 감염 예방 효과가 지속되는 기간, 고혈압 환자가 심근경색을 경험하기까지의 기간, 폐암 환자의 진단 후 사망까지의 기간 등이 있다. 이처럼 생존분석은 질병 발생과 관련된 위험 요인을 평가하고, 치료 또는 예방 조치의 효과를 측정하는 데 활용된다.

이 장에서는 생존분석의 주요 개념을 이해하고, 생존 데이터의 특징을 다룬다. 또한, 역학 연구에서 주로 사용되는 콕스 비례위험 모형(cox proportional hazard model)의 개념을 다루며, 그 원리를 이해하고 분석 결과를 올바르게 해석하는 데 중점을 둔다. 콕스 비례위험 모형은 Excel에서 분석 기능을 제공하지 않으므로, 이를 분석하기 위해 R, SPSS, SAS 등 다른 통계프로그램을 활용하는 방법을 학습할 필요가 있다.

1. 생존분석의 주요 개념

1) 생존자료의 특징

생존자료에서는 연구대상자가 특정 사건이 관측되기 전에 연구 기간이 종료되거나, 중도 탈락 등의 이유로 인해 정확한 사건 발생 시점을 관측하지 못하는 경우가 발생할 수 있다. 이러한 경우를 중도절단(censoring)이라고 한다. 예를 들어, 노출군과 비노출군을 대상으로 결과를 관찰하는 코호트연구에서 한 연구자가 연구 도중 철회 의사를 밝혀 더 이상 관찰할 수 없게 되었다고 가정하자. 이 경우, 해당 연구자의 연구 탈락 시점까지의 관찰 기간은 알 수 있지만, 최종적으로 사건(예: 질병 발생, 사망 등)이 발생한 시점은 알 수 없다. 따라서 이 연구 대상자의 데이터는 중도절단되었다고 정의할 수 있다. 생존자료는 이와 같이 중도절단이라는 특성을 가지므로, 일반적인 통계 분석 방법을 적용하기 어렵다. 따라서 생존 분석에 적합한 분석 방법을 사용하여 데이터의 특성을 반영한 올바른 해석을 수행해야 한다.

(1) 생존데이터의 표기

연구대상자 n명에 대해 생존시간을 T_1, T_2, …, T_n이라고 하고, 중도절단시간은 C_1, C_2, …, C_n이라고

하자. 실제로 관찰되는 값은 생존시간이나 중도절단시간 중 더 짧은 시간을 관찰할 수 있다. 따라서 관찰되는 값 $Y_i=\min(T_i, C_i)$로 정의할 수 있으며, 해당 시간이 사건이 발생한 시간인지를 나타내는 지시변수(indicator variable)는 다음과 같이 정의할 수 있다.

$$\delta_i=\begin{cases}1, \text{ death}\\0, \text{ censored}\end{cases}$$

즉, $\delta=1$이면 사망(사건)이 발생한 것이고, $\delta=0$이면 중도절단된 경우를 의미한다. 보통 생존분석에서 T_i와 C_i는 서로 독립이라고 가정한다.

연구대상자 n명에 대해 p개의 설명변수와 생존시간, 중도절단 여부가 수집되었다고 하자. i번째 개체의 데이터는 $(x_{1i}, x_{2i}, \cdots, x_{pi}, y_i, \delta_i)$와 같이 나타낼 수 있다.

(2) 생존함수와 위험함수

생존시간 $T(\geq 0)$는 확률밀도함수 $f(t)$와 분포함수 $F(t)$를 갖는 확률변수라고 가정하자. 이때 생존함수는 확률변수 T가 t보다 클 확률로, 사건이 t시점 이후에 발생할 확률을 의미한다.

$$S(t)=P(T>t)=1-F(t),\ 0\leq S(t)\leq 1$$

위험함수는 t시점까지 생존했을 때 t시점 바로 직후에 사건이 발생할 수간 위험률을 의미한다. 위험함수는 다음과 같이 정의되며, 생존시간의 확률밀도함수 $F(t)$와 생존함수 $S(t)$로 표현할 수 있다.

$$h(t)=\lim_{dt\to 0}\frac{P(t<T\leq t+dt\,|\,T\geq t)}{dt}=\frac{f(t)}{S(t)}$$

2. 콕스의 비례위험모형

콕스 비례위험 모형(Cox proportional hazard model, Cox PH Model)은 생존 데이터에서 시간에 따른 사건 발생 위험을 분석하기 위한 준모수적(semiparametric) 회귀 모형이다. 이 모형은 독립변수가 생존시간에 미치는 영향을 평가하면서, 생존시간의 특정한 분포(예: 지수 분포, 와이블 분포)를 가정하지 않고 분석할 수 있는 장점이 있다. 콕스 회귀 모형은 위험함수(hazard function)를 기반으로 하

며, 특정 시점 t에서 사건이 발생할 순간 위험도를 분석한다. p개의 독립변수 $x_1, \cdots, x_p$에 대해 콕스 회귀모형에서는 다음과 같이 위험함수가 정의된다.

$$h(t|x)=h_0(t)\exp(\beta_1 x_1+\cdots+\beta_p x_p)$$

여기서 $h_0(t)$는 시간 t에서 기저 위험함수(baseline hazard function)로 독립변수의 값이 모두 0일 때의 시간에 따른 위험도를 의미한다. $\beta_1, \cdots, \beta_p$는 회귀계수로 $\exp(\beta_j)$는 위험비(hazard ratio)로 해석된다.

1) 비례위험 가정

콕스 모형의 핵심 가정은 비례위험 가정(proportional hazards assumption)이다. 이는 어느 환자의 위험도와 다른 환자의 위험도의 비가 시간에 의존하지 않고 일정하다는 의미이다. i번째 환자와 j번째 환자의 위험함수의 비는 다음과 같다.

$$h_i(t)/h_j(t)=\exp\{\beta_1(x_{i1}-x_{j1})+ \cdots +\beta_k(x_{ik}-x_{jk})\}$$

즉, 위험함수의 비는 시간에 의존하지 않고 일정하다. 비례위험 가정을 확인하기 위해서 그래프를 통해 시각적으로 위험함수가 평행한지 확인한다. 이는 $\log \hat{H}(t)$ 또는 $\log(-\log \hat{S}(t))$의 그래프를 그려서 평행하게 나타나면 비례위험 가정이 만족된다고 볼 수 있다.

2) 회귀계수의 해석

단순 콕스 회귀모형에서 회귀계수 β_1에 대한 해석은 $\exp(\beta_1)$를 계산하여 위험비(hazard ratio)로 해석한다. 즉, x가 한 단위 증가할 때 사건 발생 위험이 $\exp(\beta_1)$배가 된다.

- 만약 $\exp(\beta_1)>1$ 이면, x가 한 단위 증가할 때 사건 발생 위험이 증가함을 의미하고,
- $\exp(\beta_1)<1$보다 작으면 x가 한 단위 증가할 때 사건 발생 위험이 감소하여 예방 효과가 있다고 볼 수 있다.

예를 들어, 흡연여부(1=흡연, 0=비흡연)의 회귀계수가 $\hat{\beta}_1=0.69$라면 위험비 $HR=\exp(0.69)=2.0$이다. 즉, 흡연자의 사건 발생 위험은 비흡연자의 사건 발생 위험의 2배라고 해석할 수 있다.

다중 콕스 회귀모형에서 $\exp(\beta_j)$를 위험비로 해석할 수 있다. 그러나 여러 변수가 동시에 포함되므로, 특정 변수의 회귀계수는 "다른 변수들이 일정한 상태에서" 해석해야 한다. 즉, 다중 콕스 회귀모형에서는 개별 변수의 효과를 해석할 때, 다른 변수들의 영향을 통제한 상태에서 해당 변수의 위험비(adjusted HR)로 해석해야 한다. 예를 들어, 흡연(Smoking), 나이(Age)가 생존 시간에 미치는 영향을 분석하는 경우, 콕스 회귀분석 결과 다음과 같은 식을 얻었다고 하자.

$$h(t)=h_0(t)exp(0.69\ \text{Smoking}+0.03\ \text{Age})$$

여기서 각 회귀계수에 대한 해석은 다음과 같다.

- exp(0.69)=2.0: 나이가 동일할 때 흡연자가 비흡연자보다 사건 발생 위험이 2배 높다.
- exp(0.03)=1.03: 흡연 여부가 일정한 상태에서 1세 증가할 때 사건 발생 위험이 1.03배 증가한다.

즉, 단순 콕스 회귀모형에서는 변수 하나의 영향만 해석하지만, 다중 콕스 회귀모형에서는 다른 변수들의 영향을 통제한 상태에서 해당 변수의 위험비로 해석해야 한다.

3) R을 이용한 콕스 회귀분석

다음 R code는 Cox 비례위험 회귀분석(Cox proportional hazards regression)을 수행하는 코드 형태이다. 이를 통해 특정 독립 변수(x)가 생존 시간(futime)에 미치는 영향을 분석할 수 있다.

R code	library(survival) cox <- coxph(Surv(futime, fustat = 1) ~ x, data = df) summary(cox)

콕스 회귀분석을 실행하기 위해서는 survival 패키지에 내장된 coxph( )함수를 이용한다. survival 패키지를 설치하고 라이브러리를 호출한다. coxph(Surv(futime, fustat=1)~x, data=df) 함수는 콕스 비례위험 회귀분석을 수행하는 함수이다. 여기서 종속변수는 생존시간(futime)과 사건발생여부(1=사건발생, 0=중도절단)으로 해당 데이터를 Surv( )함수를 이용하여 생존데이터로 변환한다. 이때 fustat=1은 사건발생여부에서 사건발생이 1임을 의미한다. 만약, 사건발생을 나타내는 범주가 0이라면 0으로 지정해주면 된다. 위 예제에서 독립변수는 x 하나이지만 다른 독립변수를 추가한 다중 콕스 회귀분

석을 수행하기 위해서는 $x_1+x_2+x_3$와 같이 "+"로 연결한다.

다음은 콕스 회귀분석 summary( ) 결과의 예시이다. 여기서 독립변수 x의 회귀계수는 0.045이며, p-value=0.0002이므로 생존에 유의한 영향을 미친다. 위험비(HR)는 exp(0.045)=1.046으로 x가 한 단위 증가하면 생존 위험이 1.046배가 된다. 즉, x가 증가할수록 생존 위험이 증가하며, 4.6% 증가한다.

```
coxph(formula = Surv(futime, fustat)~ x, data = df)

          coef  exp(coef)  se(coef)      z     p-value
x         0.045    1.046    0.012    3.75    0.0002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Likelihood ratio test=15.4  on 1 df,   p=8.5e-05
Wald test            =14.1  on 1 df,   p=1.9e-04
Score (logrank) test =14.7  on 1 df,   p=1.2e-04
```

콕스 비례 회귀분석도 로지스틱 회귀분석과 마찬가지로 다중 회귀분석을 수행할 수 있으며, 이를 통해 보정된 위험비(adjusted HR)를 추정할 수 있다. 다중 로지스틱 회귀분석의 예시와 R code를 활용하여 해석해보길 바란다.

참고 문헌

1. 국립보건연구원[인터넷]. 한국인유전체역학조사사업 [Cited 2025 Apr 16]. Available from: https://nih.go.kr/ko/main/contents.do?menuNo=300563.
2. 김재희. R을 이용한 생존분석 기초. 파주: 자유아카데미; 2016.
3. 박미라, 이재원. Excel을 이용한 의학데이터의 통계분석. 파주: 자유아카데미; 2024.
4. 박미라, 이재원. 생명과학연구를 위한 통계적 방법. 파주: 자유아카데미; 2006.
5. 이우주, 원성호, 김호. 보건통계학 개론. 서울: 경문사; 2024.
6. 이준영. 한눈에 알 수 있는 보건·의학 통계학. 서울: 범문에듀케이션; 2021.
7. 차재형. R과 함께하는 의학통계. 파주: 자유아카데미; 2024.
8. Agresti A. Categorical Data Analysis. 3rd ed. Hoboken, NJ: Wiley-Blackwell; 2012.
9. Glantz SA. Primer of biostatistics, seventh edition. 7th ed. New York, NY: McGraw-Hill Medical; 2011.

역학조사관 교육
표준교재

Textbook For
Korea Field Epidemiology
Training Program





2장

감시자료의 분석 역량강화







제1절 감시자료의 시계열 추세파악 및 실습
제2절 R surveillance 패키지를 이용한 유행 감지방법

집필자 천병철

제1절 감시자료의 시계열 추세파악 및 실습

1 시계열 자료의 이해

공중보건감시자료는 모두 시간에 따라서 관찰된 자료이다. 법정감염병처럼 매일매일 관찰된 자료도 있고, 인플루엔자 감시처럼 주간 감시도 있으나 이들은 모두 연속적으로 몇 년 혹은 몇 십 년간 시간에 따라서 관찰된다. 이와 같이 시간에 따라서 관찰된 자료를 시계열 자료(time-series data)라고 하며 이들은 표 2-1-1과 같은 특징을 가진다.

표 2-1-1. 시계열 자료의 주요 특성

특성	설명	예시
추세 (Trend)	시간의 경과에 따라 데이터가 증가하거나 감소하는 장기적인 경향	항생제내성의 장기적 증가, 폐렴사망률의 장기적 증가 등
계절성 (Seasonality)	일정한 주기로 반복되는 패턴을 포함하며, 계절 변화에 따른 증감이 반복되는 패턴	매년 겨울 인플루엔자의 유행, 매년 가을 쯔쯔가무시증의 유행 등
주기성 (Cyclic Behavior)	계절성보다 긴 주기의 반복적인 변동으로, 통상 2년 이상의 주기를 가짐	수두의 3년 주기 유행, 백일해의 3년 주기 유행 등
불규칙 변동 (Irregular Variation)	예측이 어려운 일회성 변동으로, 자연재해나 특정 외부 요인으로 인해 발생	수해를 입은 후 수인성감염병의 유행, 비가 온 후 말라리아 매개모기의 증가 등
자기상관성 (Autocorrelation)	특정 시점의 데이터가 이전 시점의 데이터와 상관관계를 가지는 현상으로, 미래를 예측하는 데 활용	거의 모든 시계열자료가 가지는 특성으로 이번 주 인플루엔자 발생자 수는 이전 주, 그 이전 주, 그리고 52주 전의 자료 등과 상호 연관되어 있다.
비정상성 (Non-Stationarity)	평균, 분산 등이 시간에 따라 일정하지 않은 경우를 말하며, 분석을 위해서는 정상화 작업(평균과 분산을 일정하게 만드는 작업)이 필요하다.	인플레이션
잡음 (Noise)	관측값에 포함된 무작위적이고 비규칙적인 요소. 통상 평균이 0이고 분산이 일정한 임의의 변동을 백색잡음(white noise)이라고 한다.	측정오류

여기서 시계열 자료를 이해하는 데 핵심은 자기상관성(autocorrelation)으로 이는 현재 관찰한 자료가 상호독립적이지 않고 이전 자료나 이전 연도의 같은 기간과 밀접한 연관성을 가진다는 것이다. 표 2-1-1에 있는 계절성, 주기성 등도 자기상관의 한 종류라고 할 수 있다. 예를 들어서 장출혈성대장균 감염증은 여름에 증가하고, 비브리오패혈증도 해수 온도가 높은 8월과 9월에 환자가 많이 발생한다. 인플루엔자는 겨울에 유행하는데 이러한 자료들은 계절성을 가지고 있다. 이러한 자기상관성은 시계열 자료의 분석이나 예측 모형의 정확성에 큰 영향을 미치기 때문에 이러한 자기상관성이 존재하는 경우, 이를 고려하지 않은 예측 모델은 부정확한 결과를 초래한다. 또한 시계열 자료를 이용해서 예측하고자 할 때는 추세와 비정상성 등을 같이 고려해야 하기 때문에 이들에 대한 이해는 시계열 자료를 다루는 데 필수적이다.

따라서 자료를 다룰 때 자기상관성(여기서는 시간적 자기상관성을 말하며 공간적 자기상관성을 다루지 않는다)을 확인하는 것이 중요하다. 자기상관성을 확인하는 방법은 크게 두 가지로 나눌 수 있는데 하나는 시각적으로 그래프를 그려서 확인하는 방법이다. 여기에는 자기상관 함수(ACF) 및 부분 자기상관 함수(PACF) 그래프를 주로 사용하는데, ACF는 특정 시차(lag)에서의 자기상관 계수를 보여주고, PACF는 다른 시점의 영향을 제거한 후의 자기상관 계수를 보여준다. 이 그래프들을 통해 어떤 시차에서 자기상관성이 강하게 나타나는지 시각적으로 확인할 수 있고 대부분의 통계 패키지나 R과 Python과 같은 통계 분석 언어에서는 이러한 그래프를 쉽게 그릴 수 있는 함수를 제공하고 있다.

자기상관성을 확인하는 두 번째 방법은 통계적 검정법으로 여기에 쓰이는 방법으로는 더빈-왓슨 검정(Durbin-Watson test), Ljung-Box 검정방법 등이 있다. 더빈-왓슨 검정은 회귀 분석에서 잔차의 자기상관성을 검정하는 데 사용되어 잔차의 독립성을 만족하는지를 확인하기 위한 방법이다. 더빈-왓슨 통계량은 0과 4 사이의 값을 가지며, 2에 가까울수록 자기상관성이 없음을 나타낸다. Ljung-Box 검정은 시계열 데이터의 모든 시차(lag)에 걸친 자기상관성이 존재하는지를 검정하는 방법이다. 이 외에 잔차의 고차 자기상관을 검정하는 데 사용하는 Breusch-Godfrey 검정법도 있고, ARIMA모형을 적용했을 때 AR (Auto-Regressive) 항의 유무를 확인하는 방법도 있다. R에서 이들의 사용법을 간단히 소개하면 표 2-1-2와 같다.

표 2-1-2. R을 이용한 감시자료의 시간적 자기상관성 검정법

```
# R을 이용한 acf, pacf 그래프 작성법
# 여기서는 다음과 같이 임의의 time_data라는 자료를 만들어서 이용하였다. 임의의 자료는 자료를 만들 때마다 새로
  만들어지므로 실행결과가 아래 결과와 똑같이 나오지 않음에 주의
```

```
# 임의의 time_data 생성
  > set.seed(123)
  > time_data ← arima.sim[model = list(ar = 0.7), n = 365]
# time_data의 ACF 그래프 생성
  > acf(time_data, main = "Autocorrelation Plot")
# time_data의 PACF 그래프 생성
  > pacf(time_data, main = "Partial Autocorrelation Plot")
# 더빈-왓슨 검정법: 더빈-왓슨 검정을 하는 방법이 여러 가지이지만 여기서는 lmtest 패키지를 이용한 방법을 설명한다.
  > install.packages("lmtest") # lmtest 패키지 설치 및 로드
  > library(lmtest)
  > model1 ← lm(time_data ~ 1) # 회귀 분석 수행
  > dwtest(model1) # 더빈-왓슨 검정 수행
```

```
Durbin-Watson test 수행결과
data: model1
DW = 0.72268, p-value <2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

```
# Ljung-Box 검정법
  > Box.test(time_data, type = "Ljung-Box")
```

```
Ljung-Box test 수행결과
data: time_data
X-squared = 150.02, df = 1, p-value < 2.2e-16
```

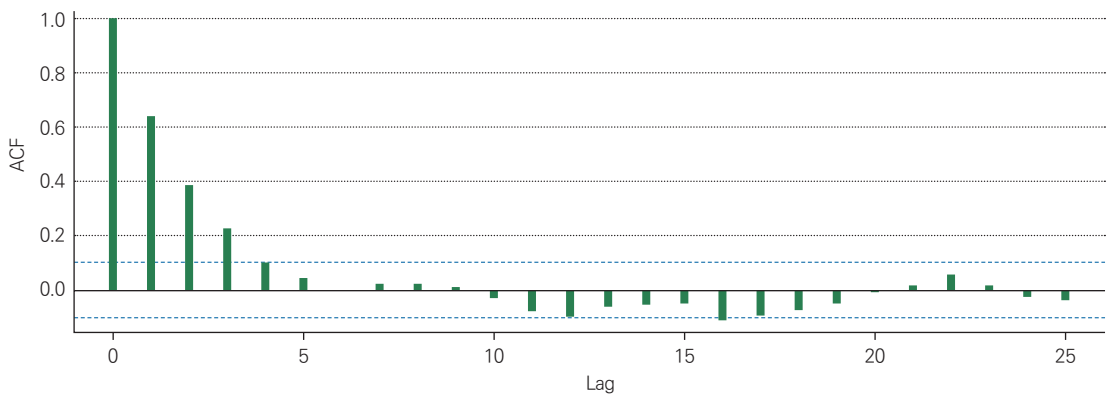


그림 2-1-1. 임의의 자료 time_data의 ACF 그래프

그림 2-1-1의 ACF 그래프의 x축은 시차를 나타내는데 시차는 현재 시점과 과거 시점 사이의 시간

간격을 의미한다. 시차 1은 바로 이전 시점의 값과의 상관관계를 나타내고, 시차 2는 2개 이전 시점의 값과의 상관관계를 표시한다. y축은 자기상관 계수를 나타내는데 자기상관 계수는 -1과 1 사이의 값을 가지며, 1에 가까울수록 양의 상관관계가 강하고, -1에 가까울수록 음의 상관관계가 강하며, 0에 가까울수록 상관관계가 약함을 의미한다. 파란색 점선의 두 줄은 신뢰구간으로 자기상관계수가 통계적으로 유의한 범위를 나타낸다. 시차 1, 2, 3이 선 밖으로 벗어나 있어서 이 시점에 유의한 자기상관이 있음을 암시한다.

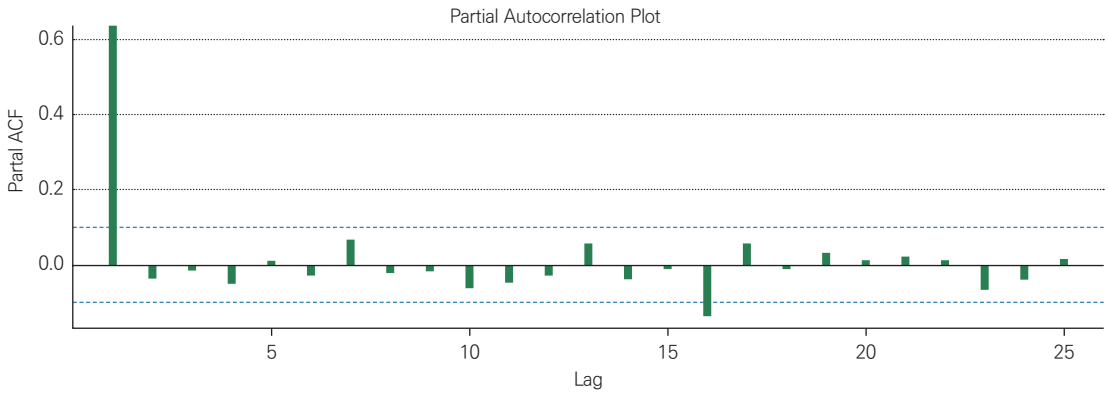


그림 2-1-2. 임의의 자료 time_data의 PACF 그래프

그림 2-1-2의 PACF 그래프는 특정 시차에서의 상관성을 중간에 있는 다른 시점의 상관성 효과를 제거한 후 측정한 것으로 특정 시점 간의 직접적인 관계를 평가하는 데 유용하다. 즉, ACF가 전체적인 상관성 여부를 측정한다면 PACF는 시차에 독립적인 순수한 상관성을 평가한다고 할 수 있다. 따라서 PACF에서는 그림 2-1-2처럼 초기 시차에서만 유의한 상관성을 보여주는 경우가 많다. ARIMA모형에서 PACF 그래프를 보고 AR(자기회귀)항의 개수를 결정하는 데 사용한다.

표 2-1-2의 더빈-왓슨 검정 결과를 보면 DW=0.72268로 2보다 많이 작고(양의 상관), p값도 매우 작아서 유의한 자기상관이 있음을 알 수 있다. 만약 DW값이 2보다 크면 음의 상관이 있음을 나타낸다. Ljung-Box 검정법에서도 p값이 매우 낮아 역시 유의한 자기상관이 있음을 보여준다.

2 시계열 자료로서의 감시자료 분석

1. 감시자료의 시계열적 자료 특성 파악

시계열 자료의 추세, 주기성, 계절성, 자기상관성을 모두 무시하고, 연보를 만들듯이 평균과 표준편차를 내고 분석하는 경우가 많은데 이것은 자료가 가지는 중요한 특성을 간과한 것이다. 통계검정이나 향후 예측 등도 시계열 자료의 특성을 반영하지 않으면 잘못되었을 가능성이 높다. 다음과 같은 감시자료를 하나 생각해 보자. 이 자료는 ECDC의 EpiSouth에서 유럽 역학조사관들을 대상으로 감시자료 분석 워크샵에 사용했던 12개 자료 중의 하나이다. 이 자료는 주별 자료로서 관찰기간 동안 지속적으로 증가하는 추세가 있으며 여름에 증가하는 뚜렷한 계절성을 보이고 있음을 우선 그래프를 통해 알 수 있다.

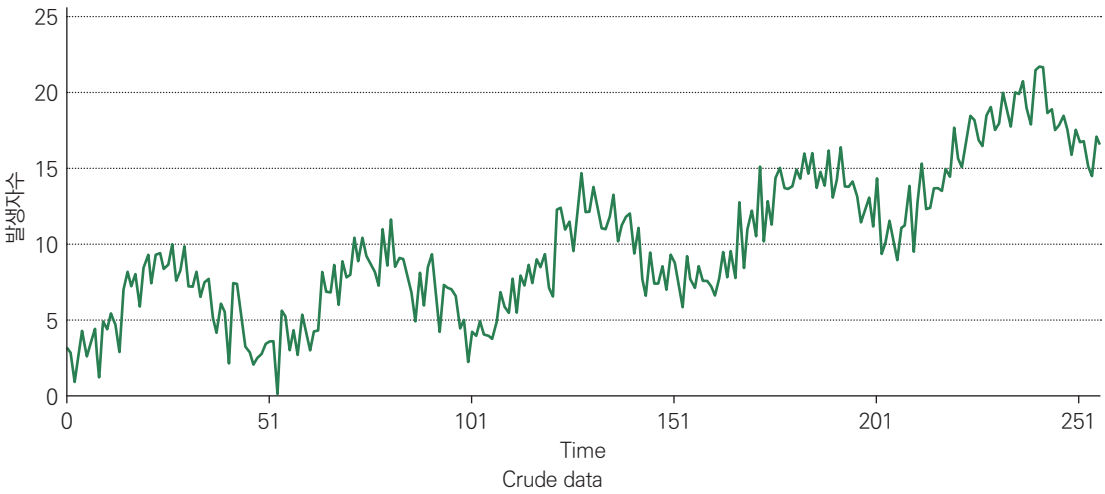


그림 2-1-3. 질병 A의 원 감시자료(raw data): 주별 발생자 수

이 자료를 분석하는 과정에서 가장 먼저 기본적인 빈도 분석도 필요하지만, 경향성에 대한 검정, 계절성에 대한 검정, 분산이 변하는지, ACF 등을 우선 점검해 보아야 한다. 이 자료로 주별 발생빈도를 그리면 그림 2-1-4와 같다. 여기서는 관찰하는 5년 동안 주별 발생자 수는 8과 9가 가장 많았다.

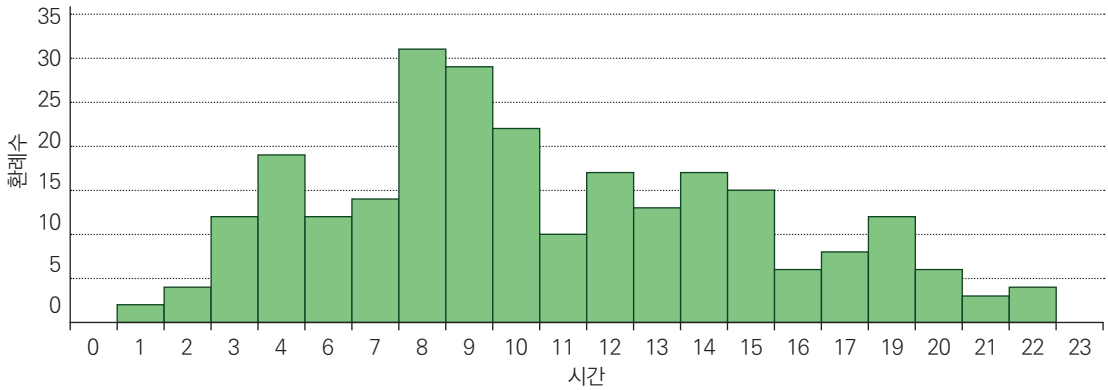


그림 2-1-4. 질병 A의 관찰기간 동안의 빈도 분포표

그림 2-1-5는 같은 자료의 추세파악을 위해서 이동평균을 먼저 그린 것이다. 이동평균 계산 시 얼마나 이동하여 평균을 낼지에 따라서(이를 window 기간이라고 한다) 원래의 그래프 모양에서 점차 직선으로 바뀌게 되는데 이 감시자료를 잘 나타내는 이동값을 선택해야 한다. 여기서는 직선 경향성을 보기 위해 이동평균의 윈도우 기간을 25주로 잡아서 계산한 것이다. 이 그래프의 경우 추세 검정에 대한 p 값은 0.001 이하로 유의한 양의 추세를 보여주고 있다.

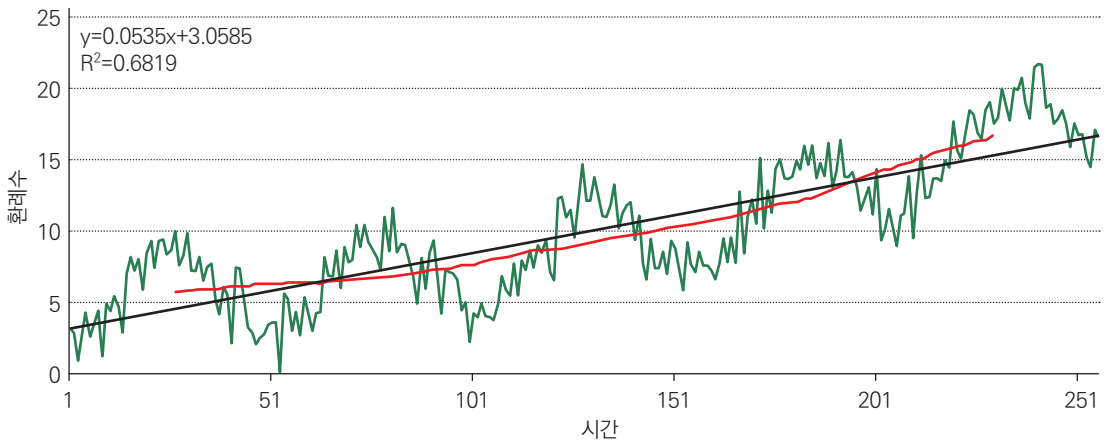


그림 2-1-5. 질병 A의 추세파악을 위한 이동평균(붉은 선)과 추세검정(검은 직선)

그림 2-1-6은 이동평균의 윈도우 기간을 16주로 잡아서 그린 것이다(붉은 선). 이렇게 그리면 이동평균선은 계절성을 잘 보여준다. 만약 이보다 더 작으면 원자료와 거의 비슷해서 매끈한 곡선이 나오지 않고, 윈도우 기간을 늘리면 점차 직선으로 변해간다. 이 이동곡선을 연도별로 그리면 그림 2-1-7과

같이 된다. 일단 연도별(계열별)로 계속 증가하고 있기 때문에 이동평균의 높이가 다르다는 것과 계열 1에서 계열 5로 갈수록 사례의 정점이 약간씩 늦어지고 있음도 관찰할 수 있다.

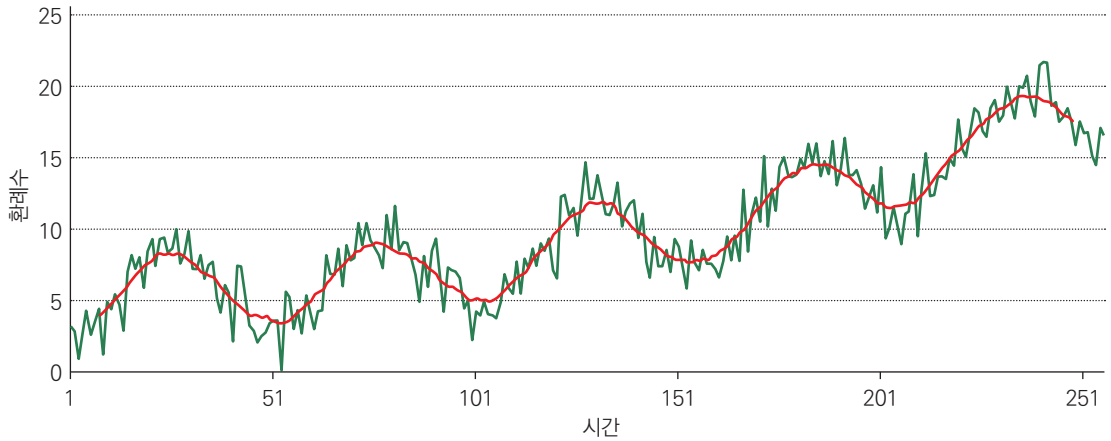


그림 2-1-6. 윈도우 기간을 16주로 설정한 이동평균(붉은 선)이 계절성을 잘 보여주고 있다

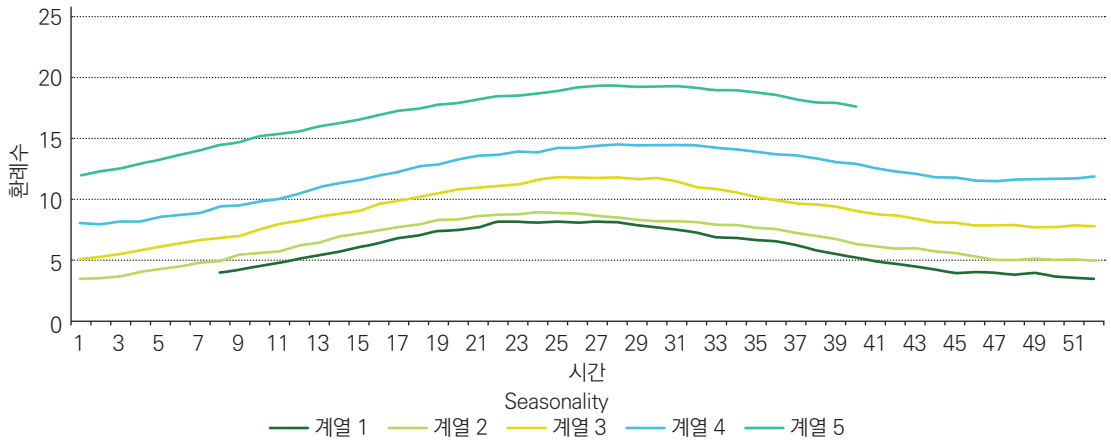


그림 2-1-7. 윈도우 기간 16주로 그린 이동평균을 연도별로 그린 것. 여기서 계열은 각 연도이다.

시계열 자료에서 시간에 따라서 평균과 분산이 변화하는지를 파악하는 것이 중요한데 이렇게 변하게 되면 비정상성(non-stationarity)의 특성을 갖는 자료가 된다. 이때는 분석을 위해서는 평균과 분산을 일정하게 만들 필요가 있게 된다. 그림 2-1-8은 해당 자료의 평균과 분산으로 평균의 변화에 따른 분산의 변화를 보기 위한 것이다. 전체적으로 평균이 커짐에 따라 분산이 약간 감소하는 듯 보이지만, 이는 통계적으로 유의하지 않았다(p=0.1010).

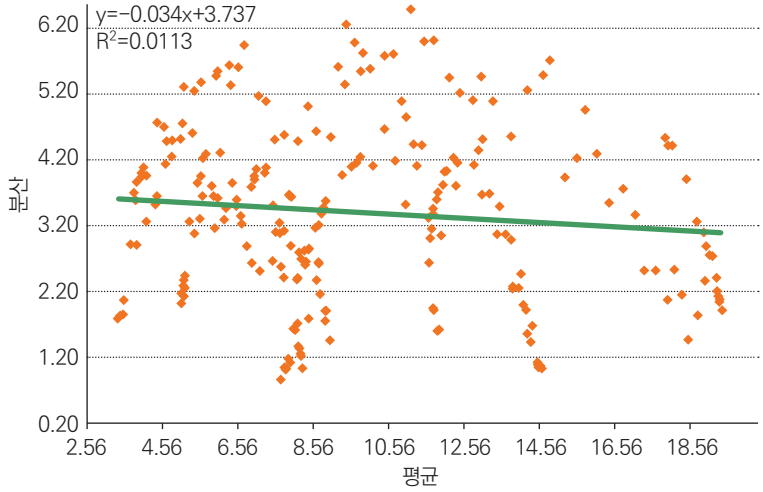


그림 2-1-8. 감시자료의 평균에 따른 분산의 변화 검정

이 자료의 자기상관성을 파악하기 위한 ACF 그래프가 그림 2-1-9이다. A는 주별 시차를 이용해서 그린 것이고 양의 방향으로 유의한 것을 알 수 있다. B는 52주 시차를 이용해서 그린 것으로 역시 52주마다 강한 상관성이 있어서 계절성이 있음을 확인할 수 있다. 이 시차는 모두 유의하였다.

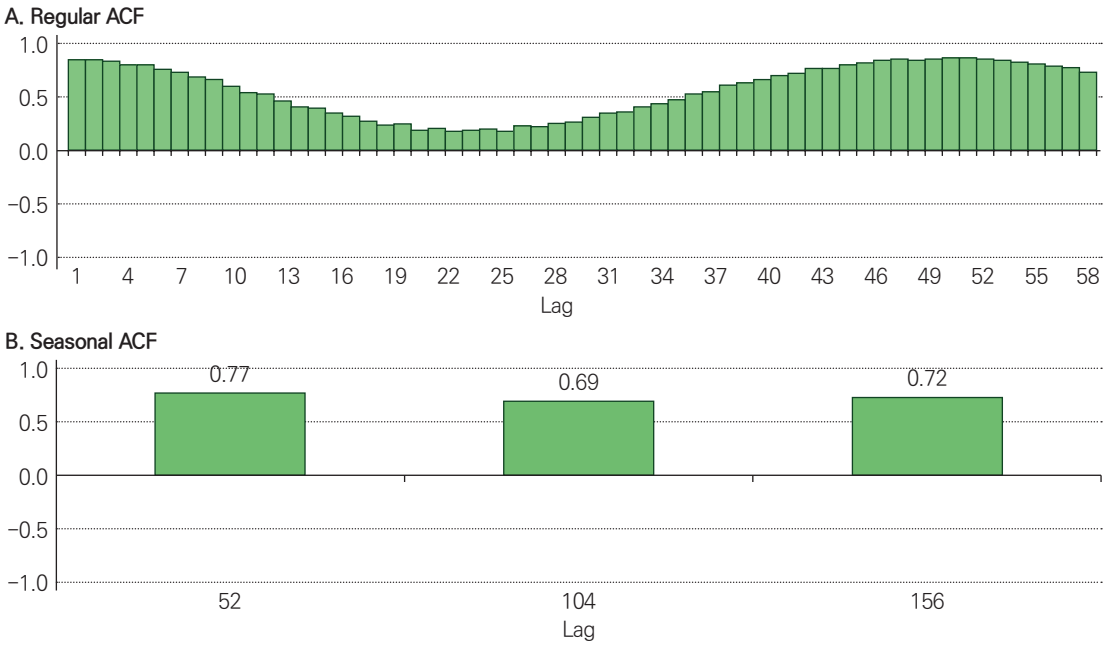


그림 2-1-9. 감시자료의 주별 ACF (A)와 계절 ACF (B) 그래프

3 감시자료의 시계열 분해

앞에서 본 바와 같이 감시자료는 추세, 계절성, 주기성 그리고 노이즈 등으로 구성되어 있으므로 이들을 분해하여 추세, 계절성, 그리고 잔차 등을 구분할 수가 있다. 이렇게 원자료를 구성하는 요소로 나누는 것을 시계열분해(decompostion)라고 한다.

표 2-1-3. R을 이용한 자료의 시계열 분해 방법

```
# 시계열분해에 사용할 자료의 준비. 여기서는 R에 내재되어 있는 자료 중 ldeaths를 이용하였는데, ldeaths는 1974년부터 1979년까지 런던의 폐질환으로 인한 사망자료이며 월별 사망자 수로 구성되어 있다.

> d_ld <- ts(ldeaths, frequency = 12)
> plot(d_ld)

# decompose 함수를 이용한 시계열 분해. 함수 decompose는 가법유형과 승법유형의 2가지 옵션이 있는데 여기서는 둘 다 사용해 보았다. 결과는 크게 차이가 없어서 가법모형으로 분해한 결과만 교재에 포함하였다.

> dec1_d_ld <- decompose(d_ld, type = "additive")
> plot(dec1_d_ld)
> dec2_d_ld <- decompose(d_ld, type = "multiplicative")
> plot(dec2_d_ld)

# stl함수를 이용한 시계열 분해

> stl_d_ld <- stl(d_ld, s.window = "periodic"")
> plot(stl_d_ld)
```

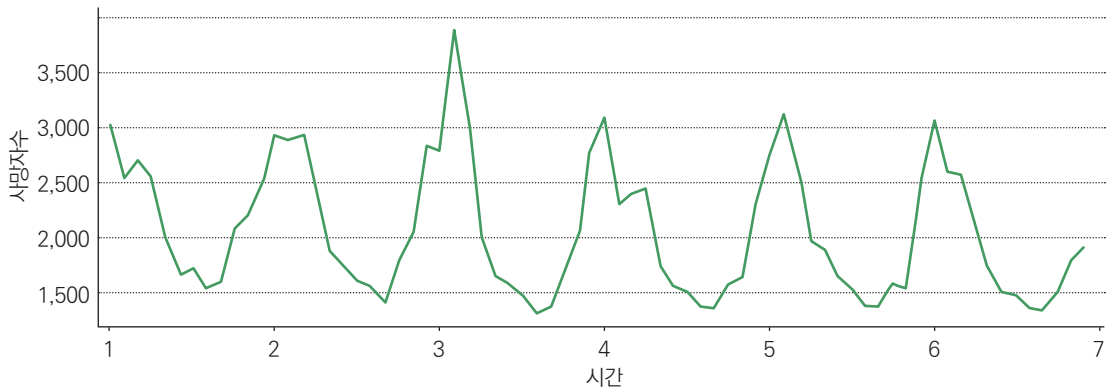


그림 2-1-10. 1974년에서 1979년까지 런던의 월별 폐질환 사망자 수

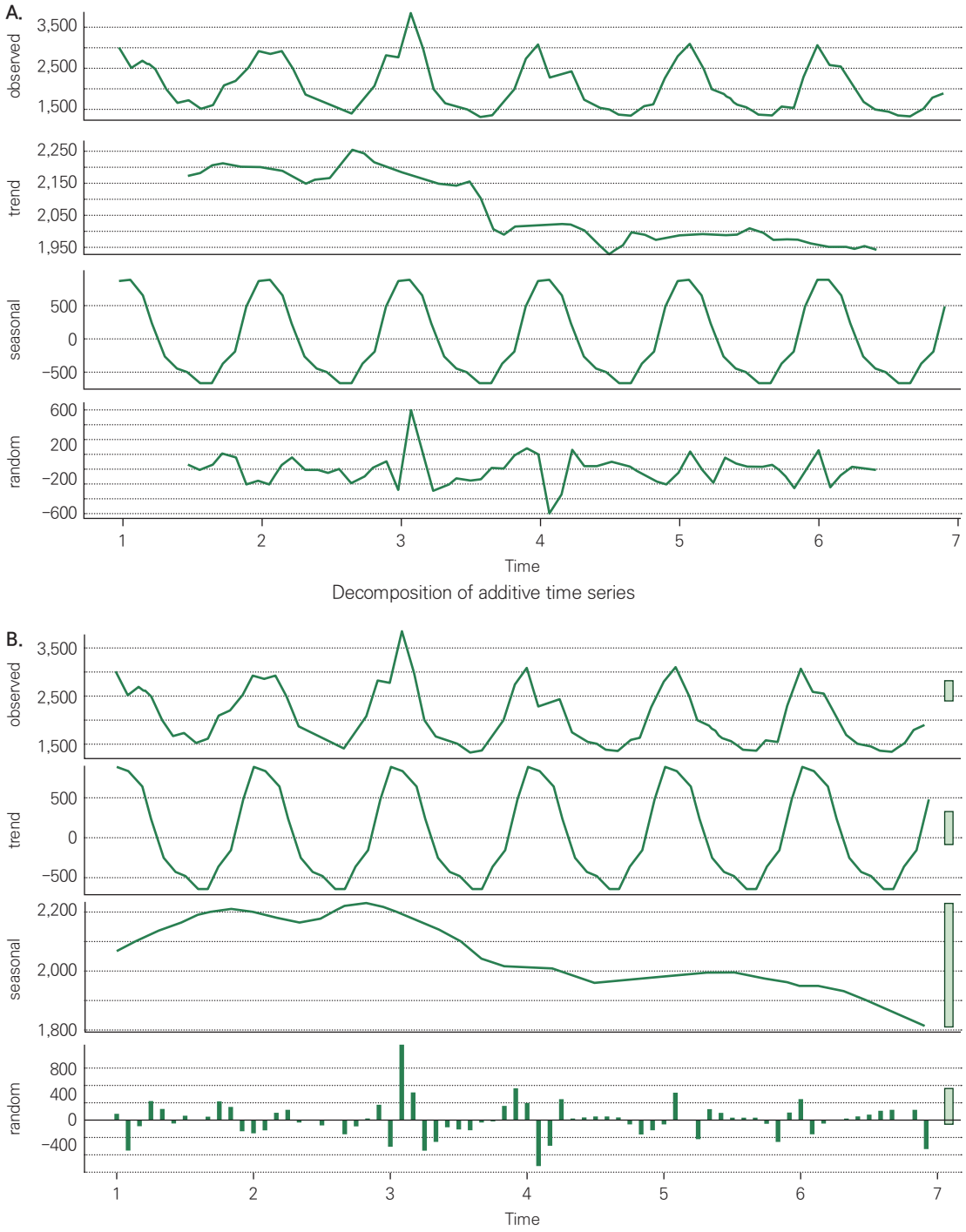


그림 2-1-11. 런던 폐질환 사망자료의 시계열 분해결과
A는 decompose 함수 사용(가법모형), B는 stl 함수 사용

표 2-1-3은 실제 자료를 시계열 분해하는 R 명령어이다. R에서 자료의 시계열 분해를 하는 함수는 여러 가지이나 여기서는 가장 흔하게 사용하는 decompose 함수와 stl 함수만 소개하고자 한다. 함수 decompose는 이동 평균(moving average)을 사용하여 추세를 추정하고, 이를 기반으로 계절성과 잔차를 계산하는 고전적인 방법이다. 반면 stl 함수는 Seasonal-Trend decomposition using Loess의 약자로, Loess 평활법을 사용하여 시계열 데이터를 추세, 계절성, 잔차로 분해한다. 이 방법은 계절성의 강도, 추세의 평활도 등을 조절할 수 있는 다양한 매개변수를 제공하여 decompose 함수보다 더 유연하고 강력하며, 복잡한 계절성 패턴을 처리할 수 있는 장점이 있다.

표 2-1-3의 R 코드를 실행시키면 우선 런던의 폐질환 사망자의 그래프를 얻을 수 있는데(그림 2-1-10) 뚜렷한 계절성을 보이는 것으로 보이지만 기타 다른 요소는 잘 확인이 안된다. 이것을 시계열분해한 것이 그림 2-1-11이다.

그림 2-1-11의 A는 decompose 함수를 이용한 것이고 B는 stl 함수를 이용한 것이다. 결과를 보면 우선 그냥 자료에서는 잘 안 보이던 경향성이 있는 것이 확인되는데 전체적으로 감소하는 경향이 뚜렷하며 특히 1976년도에 크게 감소한 경향이 있다. 이러한 특성은 원자료만 보아서는 잘 안 보이는 것이다. 계절성도 두 방법 모두 유사한 경향을 보이고 있는 것도 확인된다. 그러나 잔차의 경우는 두 방법에 약간 차이가 있는데 decompose 함수는 계산 시 앞뒤가 잘리는 반면 stl 함수는 전 과정의 잔차를 보여주고 있고, stl 함수는 각 구성요소의 스케일을 같이 제공하고 있어서 더 풍부한 정보를 보여주고 있다.

4 맺음말

여기서는 시계열자료로서의 감시자료의 특성과 시계열 분해방법을 간단히 살펴보았지만, 실제 자료분석에서는 비정상성인 자료를 정상성으로 만들기 위해서 자료를 변환하는 과정을 거치기도 하고, 평균에 따라서 분산이 증가하는 경우도 많아서 이들을 본격적인 분석 이전에 고려해야 한다. 또한 자료를 분해하다 보면, 눈으로는 전혀 계절성이나 주기성이 없어 보이던 자료가 분해에서는 뚜렷하게 나타나는 경우도 있다. 또한 장기간의 자료인 경우 이 계절성의 변화가 있는 것을 관찰하고 이를 분석할 수도 있는데 예를 들면, 쯔쯔가무시증의 장기간 감시자료를 이용해서 쯔쯔가무시증의 계절성이 약화되고 있는 것을 밝혀내기도 했다.

감시자료의 가장 중요한 목적 중의 하나는 유행을 발견해 내는 것인데, 이것이 유행인지 아닌지는 이러한 추세, 계절성, 자기상관을 모두 고려한 다음에 파악이 가능하다. 이것은 이 자료를 이용한 향후 예

측에서도 마찬가지이다. 특히 감염병의 경우 계절성이 아주 뚜렷한 것이 많다. 그런데 이러한 특성을 무시하고 연간 평균과 표준편차를 계산하여 비교하고 있는데 이는 감시자료를 대표한다고 보기 어렵다. 예를 들어서 폐구균예방접종이 국가예방접종에 포함된 이후 폐구균으로 인한 폐렴이 줄었는지를 알아본다고 하자. 폐구균폐렴은 뚜렷한 계절성이 있는 데다가 국가예방접종사업 전에도 감소하고 있었거나 혹은 증가하고 있었던 추세를 가지고 있다. 그러므로 국가예방접종사업 후에 예상되는 폐구균폐렴 환자 수는 이들을 모두 고려해야 하는 것이다. 실제로 이러한 분석에서 ARIMA모형을 이용한 단절적 시계열분석 방법을 사용하였다. 이런 예측자료를 기반으로 하는 시계열 자료의 분석에서 시계열자료의 특성을 무시한 전후 평균의 비교는 넌센스에 가깝다.

이 1절은 감시자료의 시계열적 특성을 이해하기 위한 것이고, 실제 자료에서 유행을 발견하는 방법은 2절에서 다룬다.

참고 문헌

1. 천병철. Outbreak detection methods in surveillance. 고려대학교 보건대학원 계량역학 강의록. 2024.
2. 천병철. R을 이용한 감시자료의 분석. 우즈베키스탄 감염병감시국 강의자료. 2022.
3. 천병철. 질병 감시자료의 분석 및 해석 실습. 고려대학교 보건대학원 공중보건정보학 강의록. 2020.
4. ECDC [Internet]. EpiSouth Training Module: Introduction to the analysis of surveillance data [Cited 2025 Apr 16]. Available from: http://www.episouth.org/outputs/wp5/f_t_module/index.html.

제2절 R surveillance 패키지를 이용한 유행 감지방법

1 R surveillance 패키지

1. R surveillance 패키지의 주요 분석유형과 기능

역학분야에서 가장 많이 사용되는 통계프로그램은 R이라고 할 수 있는데 이는 무료라는 장점 이외에도 역학분야의 특수한 목적에 맞게 개발되어 있는 수많은 패키지를 활용하면 복잡한 분석도 쉽게 할 수 있다는 매우 큰 장점이 있기 때문이다. 이 중의 하나인 surveillance 패키지는 감시자료를 분석하는 데 특화된 패키지로 이를 이용해서 다양한 분석이 가능하다. 감시자료는 시계열 자료이기 때문에 이 특성에 맞는 분석법이 필요한데 R의 surveillance 패키지는 시계열 자료의 특성을 기반으로 감염병 감시, 유행 감지, 예측 등 다양한 분석을 수행할 수 있다. 이 패키지를 활용하여 할 수 있는 주요 분석들은 표 2-2-1과 같다.

표 2-2-1에서 보는 것처럼 surveillance 패키지는 감시자료의 시계열 분석, 공간분석, 위험도 평가, 감염재생산수 추정과 방역조치의 평가, 지도 작성, 보고서 자동 작성 기능 등 막강한 기능을 가지고 있고, 이 기능들은 역학조사관의 업무와 상당히 일치한다. 뿐만 아니라 이 패키지를 사용하면 감염병 확산 추세를 실시간으로 파악하고, 효과적인 방역 전략을 수립하는 데 큰 도움이 될 수 있기 때문에 이 패키지 하나만 잘 익혀 두어도 역학조사관 업무의 효율이 크게 증가할 것이다. 또 surveillance 패키지는 감염병 이외의 환경질환 등 다양한 분야의 시계열 자료 분석에도 활용될 수 있는 장점이 있다. R에는 감시자료분석과 관련해서 surveillance 패키지 이외에도 rsatscan, vetsyn과 같은 유용한 패키지들이 다양하게 있으니 관심을 갖고 익힐 필요가 있다.

표 2-2-1. R의 surveillance 패키지의 분석유형에 따른 주요 기능

분석 유형	주요 기능	활용 분야
감염병 감시 및 이상 감지	• 감염병 발생률, 환자 수 등 시계열 자료를 분석하여 추세, 계절성, 주기성 등을 파악 • CUSUM, EARS, Farrington 알고리듬 등 다양한 통계적 방법을 사용하여 예상치 못한 이상 증가(유행)를 감지 • 시공간 자료 분석: 감염병 발생 지역 및 발생 시점을 함께 고려하여 시공간 클러스터를 분석하고, 이상 발생 지역을 탐지 • GIS 데이터를 활용하여 감염병 발생 분포를 시각화하고, 공간적 위험도를 분석	감염병 확산 감시, 이상 발생 탐지, 방역 전략 수립
감염병 예측 (시계열 예측과 위험도 예측)	• ARIMA, SIR 모델 등 다양한 시계열 예측 모델을 사용하여 감염병 발생 추이를 예측 • 예측 결과를 기반으로 의료 자원 확보, 방역 전략 수립 등 선제적 대응을 지원 • 환경 요인, 인구 이동 등 다양한 요인을 고려하여 감염병 발생 위험도를 예측	감염병 발생 추이 예측, 의료 자원 확보, 선제적 대응, 위험 지역에 대한 집중 감시, 예방 캠페인 등 맞춤형 방역 전략 수립을 지원
감염재생산수 추정	• 감염병의 확산 속도를 나타내는 유효 감염재생 수(Rt)를 시간 경과에 따라 추정 • Rt 추정 결과를 기반으로 방역 조치의 효과를 분석	감염병 확산 속도 평가, 방역 정책 결정 지원
데이터 시각화 및 보고	• 시계열 그래프, 공간 분포 지도, 위험도 지도 등 다양한 시각화 도구를 제공 • 분석 결과를 기반으로 보고서를 자동 생성하는 기능	분석 결과 시각화, 의사 결정 지원, 정보 공유

2 R surveillance 패키지의 기본 사용법

표 2-2-2. R surveillance package의 주요 기능별 사용코드 요약

1. 패키지 설치 및 로드

```
> install.packages("surveillance")        # 패키지 설치
> library(surveillance)                   # 패키지 로드
```

2. 데이터 준비

surveillance 패키지는 시계열 자료를 입력으로 사용한다. 데이터는 날짜, 감염병 발생 건수 등 시간 순서대로 정렬되어 있어야 한다.

```
> data(ha)                                # A형 간염 데이터셋 로드
> head(ha)                                # A형 간염 데이터 확인
```

3. 감시분석 알고리듬 적용
surveillance 패키지는 CUSUM, EARS, Farrington 등 다양한 감시분석 알고리듬을 제공하기 때문에 원하는 알고리듬을 선택하고 데이터를 입력하여 감시 결과를 얻을 수 있다(각 감시분석 알고리듬에 대해서는 참고문헌 참조).
> ha_farrington <- farringtonFlexible(measles) # Farrington 알고리듬을 적용하여 유행 발견 > ha_farrington # 결과 확인
4. 결과 시각화
surveillance 패키지는 감시 결과를 시각화하는 다양한 기능이 있으므로 감염병 발생 추이, 이상 발생 시점 등 원하는 결과를 그래프로 확인가능
> plot(ha_farrington) # 결과 그래프 출력
5. 시공간 데이터 분석
감염병 발생 지역 및 발생 시점을 함께 고려하여 시공간적 클러스터를 분석 epilink 함수 등을 사용하여 시공간적 이상 발생 지역을 탐지
공간-시간 데이터 예시 (가상 데이터) > data_spat_temp <- data.frame(> date = rep(seq(as.Date("2020-01-01"), as.Date("2020-12-31"), by = "week"), 5), > region = rep(c("A", "B", "C", "D", "E"), each = 53), > cases = sample(0:10, 265, replace = TRUE) >)
공간-시간적 이상 발생 탐지 > result_spat_temp <- epilink(data_spat_temp$cases, data_spat_temp$date, data_spat_temp$region)
6. 감염재생산수 추정
surveillance 패키지는 시간에 따른 감염 재생산 수(Rt)를 추정하는 기능을 제공 estimateR 함수 등을 사용하여 감염병 확산 속도를 평가
감염 재생산 수 추정 > result_rt <- estimateR(measles$incidence, method = "parametric_si", config = list(mean_si = 2.8, std_si = 1.5))

기본적인 surveillance 패키지의 주요 기능에 따른 사용코드를 간추리면 표 2-2-2와 같다. 붉은색이 R 명령어 코드이고, #은 이에 대한 설명이다.

3 R surveillance 패키지를 이용한 유행 발견의 알고리듬

감시의 가장 중요한 목적 중의 하나는 유행(outbreak)을 발견하기 위한 것이다. 유행은 평소 기대했던 수보다 많이 발생하는 것을 의미한다. 평소 기대했던 수보다 많다는 의미는 통상 평상시 발생범위의

95%신뢰구간을 벗어나는 발생수준을 말한다. 감시자료의 분석에서는 이상치(aberration)라는 용어도 많이 사용하는데 이는 아직 유행인지 아닌지는 모르지만 정상범위로 보기 어려운 발생 수를 의미한다. 즉, 이상치는 유행을 발견하기 위한 전초단계로 실제 감시자료 분석에서 유행을 하나라도 놓치지 않으려면, 즉 유행발견의 민감도를 높이려면 이상치들을 발견해 내야 한다. 따라서 감시자료의 유행의 발견에서는 이상치를 검출하고자 하는 매우 다양한 알고리듬들이 개발되어 있다.

R surveillance 패키지에도 시계열 자료의 이상치(aberrration)를 탐지하기 위하여 다양한 알고리듬을 가지고 있는데 대표적인 알고리듬으로는 다음과 같은 것들이 있다.

1. 패링턴 알고리듬

패링턴(Farrington)이 개발한 알고리듬으로 고전적으로 오래전부터 사용해 오던 방법이다. 과거 발생자료를 기반으로 예상 발생 범위를 계산하고, 실제 발생 건수가 이 범위를 벗어나는 경우 이상치로 판단하는 기본 원리가 잘 반영되어 있다. Farrington 알고리듬은 과거 5년 동안의 발생자료를 사용하여 (참고: 감염병의 경우 보통 과거 5년 동안의 자료를 사용) 현재 주(week) t에 대한 기댓값 $\hat{\lambda}_t$와 신뢰구간을 설정한다.

패링턴 알고리듬은 과거 데이터의 이동 윈도우(moving window)에서 일반화 선형 모형(Generalized Linear Model, GLM)을 사용하여 현재 시점의 기대값을 계산하는데, 식으로 표현하면 식 1과 같다.

$$\text{식 1: } \hat{\lambda}_t = \exp\left(\beta_0 + \sum_{i=1}^{n} \beta_i X_i\right)$$

여기서 $\hat{\lambda}_t$는 시간 t에서의 예상사례 수, X_1는 과거 b년 동안의 같은 주(week)의 발생자 수이다. 이때 분산은 감마분포나 과산포 포아송모델(overdispersed Poisson model)을 이용하여 추정한다(식 2).

$$\text{식 2: } \hat{\sigma}_t^2 = \hat{\phi}\hat{\lambda}_t$$

식 2에서 $\hat{\phi}$는 과산포계수(overdispersion factor)이다. 포아송모형이므로 분산과 기댓값이 같지만 과분산계수로 보정하여 분산을 계산한 것이다. 이와 같이 식 1의 기댓값과 식 2의 분산에서 추정한 표준편차를 이용해서 기댓값의 95%신뢰구간을 구할 수 있는데 이것이 유행기준이 된다.

R surveillance package에서는 farringtonFlexible ( ) 함수를 통해 유행 여부를 구할 수 있다.

2. 누적합 알고리듬

누적합(Cumulative Sum, CUSUM) 알고리듬은 유행을 좀 더 민감하게 발견하기 위하여 개발된 것으로 증후군 감시나 생물테러 등 유행을 빨리 발견해 내야 하는 감시의 분석방법에 자주 사용되고 있다. 누적 합을 사용하여 작은 변화도 감지할 수 있는 알고리듬이며, 지속적인 이상 발생을 탐지하는 데 유용하고, 감염병 유행의 시작 또는 지속적인 증가를 감지하는 데 사용된다. 원리는 각 관찰값과 기준값(평균)의 차이를 누적하여, 이 누적합이 특정 임곗값을 초과하면 이상치로 감지하는 것이다. 이해가 쉽고 계산이 간단한 장점이 있다.

CUSUM의 누적합은 다음과 같이 정의된다.

$$식\ 3: S_t=\max(0, S_{t-1}+(X_t-\mu_0-K))$$

여기서 S_t는 시점 t에서의 CUSUM 값, S_{t-1}은 이전 시점의 CUSUM 값, X_t는 시점 t에서 관측된 값, μ_0는 기준값(평균값, 즉 정상 상태의 기댓값)이며, k는 감지 민감도를 조절하는 상수이다(보통 k=h/2). 이때 CUSUM의 초기값은 0에서 시작한다. 즉 $S_0=0$.

CUSUM 값이 사전에 설정된 임곗값 h를 초과하면 이상 발생으로 판단한다. 식으로 표시하면 식 4와 같다.

$$식\ 4: S_t>h$$

여기서 임곗값 h가 작으면 빠르게 탐지하는 장점이 있지만 이상치가 아닌 것도 잘못 탐지하여 잘못된 알람(false alarm)이 생기며, h가 크면 신뢰도는 높아지지만 민감도가 낮아지기 때문에 h값의 설정이 중요하다. 이와 같이 CUSUM 알고리듬은 과거의 작은 변화가 누적되어 미세한 변화도 감지 가능하고, 과거 자료와 비교하여 변화점을 감지하는 순서 기반 접근 방식을 사용하는데, k와 h의 설정에 따라 탐지 성능이 달라지는 특성을 가진다.

R surveillance에서는 cusum ( ) 함수를 통해 사용한다.

3. 조기이상치보고체계 알고리듬

조기이상치보고체계(Early Aberration Reporting System, EARS)는 미국 CDC에서 개발한 알고리듬

으로, 단기적인 이상 발생을 신속하게 감지하는 데 사용하며, 과거 발생자료를 기반으로 예상 발생 범위를 계산하고, 실제 발생 건수가 이 범위를 벗어나는 경우 이상치로 판단하는 방법이다. EARS는 최근 데이터(단기 창, short-term window)와 과거 데이터(장기 창, baseline window)를 비교하여 감염병 발생 이상 여부를 판단한다. 즉, 최근 며칠(예: 3일)의 발생 건수를 평균 및 표준편차와 비교하여 Z-Score를 계산하고 특정 임곗값을 초과하면 이상 탐지(Alert)를 발생시키는 것이다.

이때 사용하는 EARS 알고리듬은 과거 자료의 사용방식에 따라 C1, C2, C3의 3가지가 있다.

표 2-2-3. 조기이상치보고체계 알고리듬의 임곗값

EARS 알고리듬	감시 방법	특징
C1 (Fixed Baseline)	단순 이동 평균과 비교	단순 이동 평균과 비교하므로 민감도가 가장 낮으며 주요 이상만 탐지할 때 적합
C2 (Moving Baseline)	이동 평균과 표준편차 활용	평균과 표준편차를 같이 활용하여 최근 W일 간의 변동성을 반영하여 감시함으로써 안정적인 탐지 가능
C3 (Weighted Moving Baseline)	이동 평균과 가중치를 활용	최근 데이터에 가중치를 줌. 민감도 높으며 갑작스러운 변화에 더욱 민감하게 반응

C1 방법은 최근 W일 동안의 평균을 기준으로 하여 현재 값과 비교하는 방식이다. 예를 들어 최근 7일간의 발생자 수 평균과 오늘의 발생자 수를 비교하여 2 또는 3 이상의 차이가 발생하면 이상으로 간주한다. C2 방법은 이동평균과 표준편차를 이용하여 95% CI를 구하는 것으로 안정적인 탐지가 가능하다. C3 방법은 최근 자료일수록 높은 가중치를 주어서 가중이동평균을 계산한 것으로 갑작스러운 변화에 민감하게 반응하도록 하여 민감도를 높인 것이다(각 수식은 참고문헌 참조).

R surveillance는 EARS C1 ( ), EARS C2 ( ), EARS C3 ( ) 함수를 이용해 구한다.

4. 일반화우도비 알고리듬

일반화우도비(Generalized Likelihood Ratio, GLR) 알고리듬은 일반화우도비검정법을 사용하여 이상치를 탐지하는 알고리듬으로, 다양한 분포를 따르는 자료에 적용할 수 있으며, 감염병 발생 패턴의 변화를 감지하는 데 사용할 수 있다. 특징으로는 조기 감지에 적합하여 유행 발생 직후에 유행을 식별하여 대응 시간을 단축할 수 있으며, 포아송, 음이항 분포 또는 기타 분포를 따르는 다양한 데이터 유형에 대한 적응성이 뛰어나고, 연속 모니터링에 적합하여 자동 감시 시스템이나 실시간 감시에 유용하다는 것이다. 또한 임곗값을 조정하여 거짓 경보율을 제어할 수 있으며 고정 임곗값 접근 방식과 달리

GLR은 과거 추세와 가능성 변화를 모두 고려한 견고한 방법이다.

R surveillance에서는 glr ( ) 함수를 통해 사용할 수 있다.

5. OutbreakP 알고리듬

포아송 분포를 기반으로 이상치를 탐지하는 알고리듬으로 감염병 발생 건수가 포아송 분포를 따르는 경우에 적합하며, 감염병 유행의 시작을 탐지하는 데 사용된다. 즉, 실제 발생 건수와 예상발생 건수를 포아송 분포를 사용하여 비교하고 이때 p값을 계산하여 p가 0.05보다 작으면 해당 시점을 이상치 발생으로 판단하는 방법이다.

R surveillance에서는 outbreakP ( ) 함수를 이용해 사용할 수 있다.

6. CDC 알고리듬

CDC 알고리듬은 가장 고전적인 방법으로 과거 b년(보통 5년)간의 자료를 이용해서 평균과 표준편차를 구한 후 이를 이용해서 95% 신뢰구간을 구해(즉, 2 표준편차) 이를 유행기준으로 정하는 방식이다.

4 R surveillance 패키지를 이용한 유행 발견 실습

R surveillance 패키지에 내재되어 있는 자료를 이용해서 유행 발견 실습을 해보자. Surveillance 패키지에는 30개가 넘는 감시자료가 들어있다. 참고로 R에서 surveillance 패키지에 포함된 데이터셋을 보는 명령어는 > data (package="surveillance")이다. 이 중에서 독일에서 1990년에서 1995년 *Salmonella agona* 유행자료를 불러온 후 박스 2-2-1을 실행해 보자.

박스 2-2-1. R surveillance 패키지의 패링턴 알고리듬을 이용한 유행의 발견

```
> install.packages("surveillance")                    # 패키지 설치
> library(surveillance)                               # 패키지 로드
```

```
> data("salmonella.agona")                                  # 데이터 이용
> summary(salmonella.agona)                                 # 312입력자료 확인(결과는 생략)
> plot(salmonella.agona)                                    # 1990~1995년간 발생자료 확인(그림 2-2-1)

# 패링턴방법을 적용하려면 자료가 "sts" (surveillance time series) 객체여야 하므로 먼저 현재의 자료(disProg
  객체)를 sts로 바꾸어 준다.
> salmonella_sts <- disProg2sts(salmonella.agona)           # disProg 객체를 sts 객체로 변환
> farrington_result <- farringtonFlexible(salmonella_sts)   # Farrington 알고리듬 적용
> plot(farrington_result)                                   # 그래프로 결과 확인(그림 2-2-2)
```

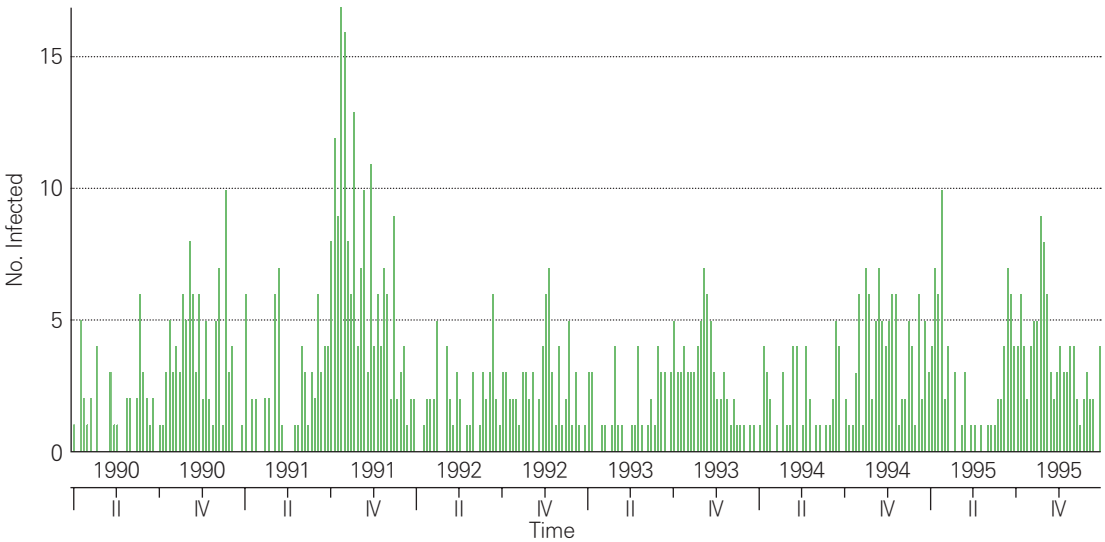


그림 2-2-1. 독일의 Salmonella.agona 발생자 수(1990~1995)

박스 2-2-1에서 우선 자료를 살펴보면 1990년에서 1995년까지 발생자료를 볼 수 있다(그림 2-2-1). 이 자료에 패링턴 알고리듬을 적용하기 위해서 sts 객체로 전환한 후 패링턴 알고리듬으로 이상치를 감지한 것이 그림 2-2-2이다. 여기서 파란 선이 패링턴 알고리듬을 적용한 유행기준이고 이 유행기준보다 더 많이 발생한 경우는 x축에 붉은색 삼각형으로 표시가 되어 있다. 즉 1995년 동안에 3번의 유행이 있었다고 해석할 수 있다.

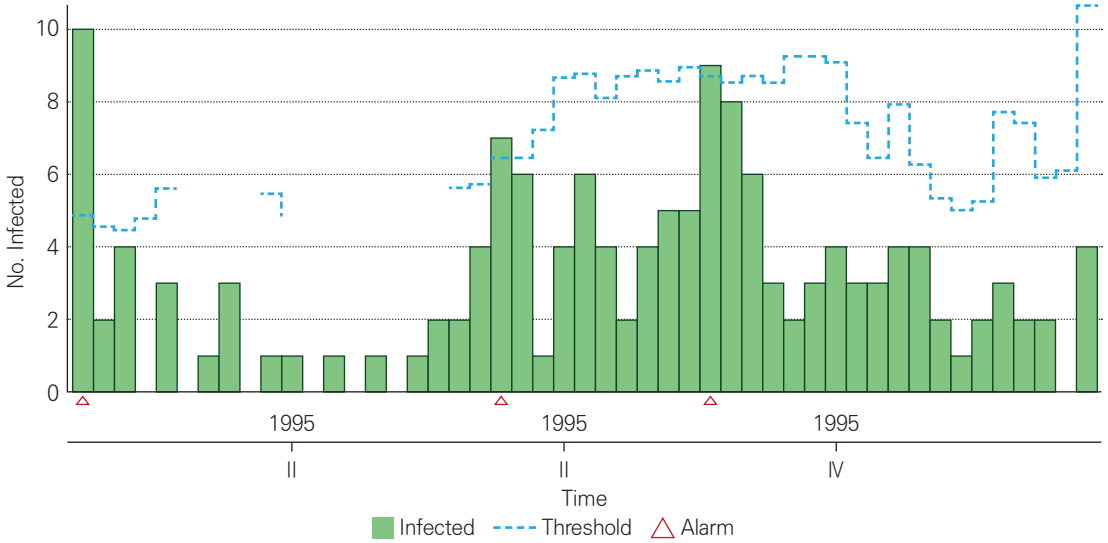


그림 2-2-2. 독일의 Salmonella.agona 감시자료에서 패링턴 알고리듬으로 감지한 유행(붉은색 삼각형)

패링턴 알고리듬을 적용할 때 control 옵션을 이용해서 설정을 조정할 수 있다(박스 2-2-2). 이들 옵션을 조절해서 유행 기준설정에 활용할 수 있다.

박스 2-2-2. 패링턴 알고리듬에서 control 옵션에 대한 설명

```
> farrington_result ← farringtonFlexible(
>  salmonella.agona,                          # 데이터셋
>  control = list(
>    range = 50:100,                          # 예측에 사용할 과거 주(week)
>    b = 5,                                   # 과거 몇 년간의 데이터를 사용할지 설정
>    w = 3,                                   # 과거에서 몇 주 동안 데이터를 평균 낼지 설정
>    alpha = 0.05                             # 유의 수준(95% 신뢰구간)
>  )
> )
```

이 *Salmonella agona* 자료는 5년간의 자료이므로 5년간의 과거 자료를 필요로 하는 방법을 사용하기는 어렵지만 GLR 알고리듬은 적용 가능하다.

박스 2-2-3. GLR 알고리듬을 적용한 유행의 발견

```
# 패키지 설치와 data salmonella.agona의 이용은 앞과 같다.
# GLR 알고리듬의 적용 – control parameters의 결정
> glr_control <- list(
> range = 1:length(salmonella.agona$observed),        # Range of observations to consider
> m = 1,                                              # Minimal cases to start monitoring
> Mtilde = 10,                                        # Maximum number of past time points for reference
> alpha = 0.05,                                       # Significance level for outbreak detection
> mu0 = NULL,                                         # Baseline mean (NULL means automatically estimated)
> change = "intercept",                               # Model change point (intercept shift in mean)
> theta = NULL                                        # Dispersion parameter (NULL for automatic estimation)
)

> glr_result <- algo.glrnb(salmonella.agona, control = glr_control)     # GLR 알고리듬의 적용
> plot(glr_result, main = "GLR Surveillance for Salmonella Agona"")     # GLR 결과 도표화
```

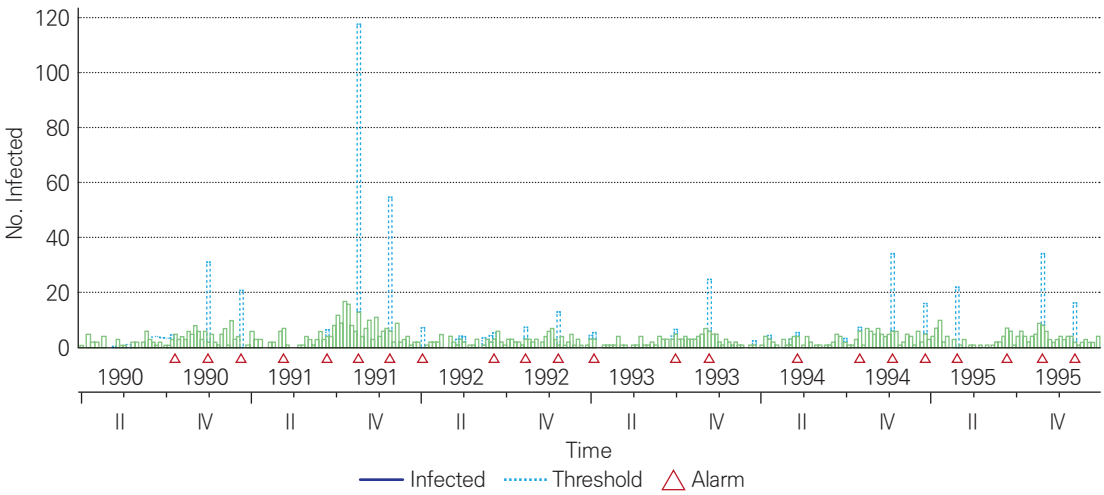


그림 2-2-3. 독일의 Salmonella.agona 감시자료에서 GLR 알고리듬으로 감지한 유행(붉은색 삼각형)

박스 2-2-2의 결과 그림 2-2-3에서 보는 것처럼 GLR 알고리듬은 1990년 중반에서부터 1995년까지 여러 번의 유행을 찾아낼 수 있는데 거의 실시간으로 유행을 파악하는 것이 가능한 장점이 있다.

이번에는 2001년 1주부터 2006년 30주까지(총 290주) 베를린 18세 이상 성인 남성의 A형 간염의 감시자료를 가지고 분석해 보자. 이 자료는 베를린 도시 내 12개 지역으로 구성되어 있지만 각 지역마다 주별 1명 내지 2명씩의 발생자료여서(그림 2-2-4), 이것을 모두 합쳐서 베를린의 자료로 만들어서 분석을 하는 것이 좋다.

박스 2-2-4. 베를린 A형간염 감시자료를 이용한 분석: 자료합치기

```
> library (surveillance)
> data (ha)                                    # 베를린의 A형 간염 감시자료 이용
> Summary (ha)
> Plot (ha)
# 베를린의 A형 간염 감시자료가 12개 지역으로 구성되어 있음을 확인. 그러나 그래프가 너무 복잡하게 나오기 때문에 여기서는
plot(ha.sts)를 이용해서 지역별로 나누어져 나온 그래프(그림 2-3-4)를 제시함.
```

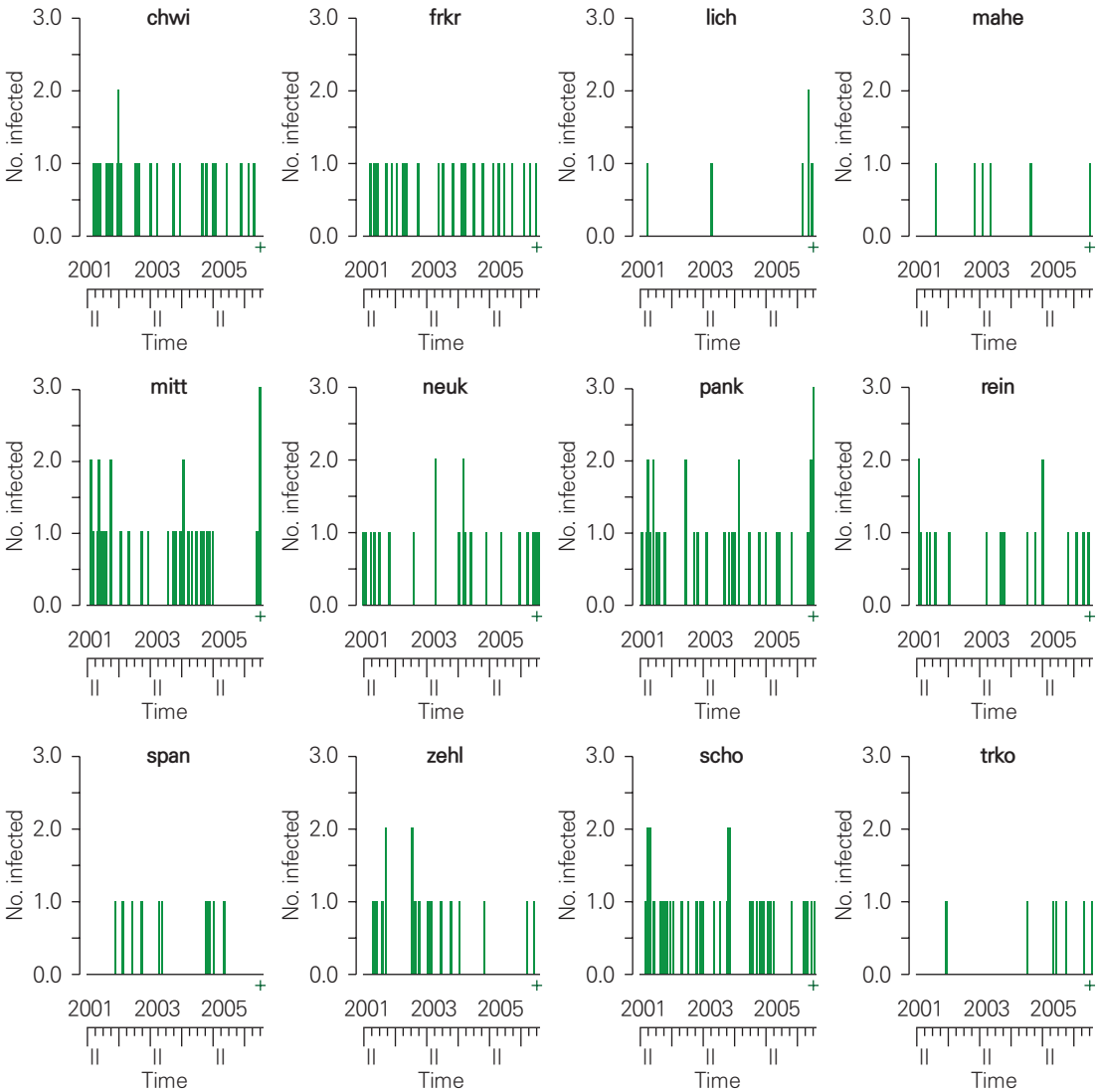


그림 2-2-4. 베를린 12개 지역의 A형 간염 감시자료(2001~2006)

```
> plot(aggregate(ha), main = "Hepatitis A in Berlin 2001-2006")
# 지역으로 나누어져 있는 것을 합하고 그래프의 제목을 지정해 줌.
```

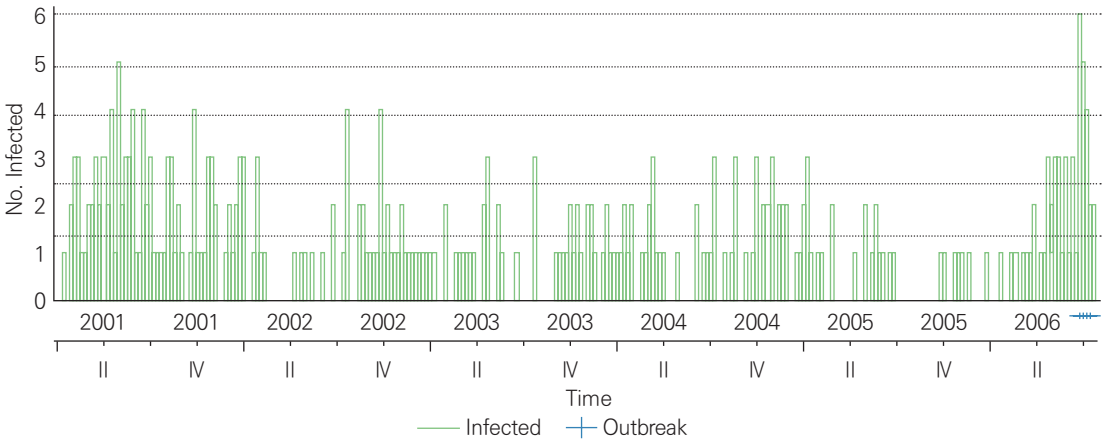


그림 2-2-5. 베를린 12개 지역을 하나로 합친 A형 간염 감시자료(2001~2006)

그림 2-2-5를 보면 12개 지역의 감사지료를 합쳤을 때 2006년 3/4분기에서 유행이 있음을 알 수 있다. 그런데 이것을 좀 더 빨리 알아내는 방법은 없을까?

이제 이 자료를 이용해서 CDC, 패링턴, 베이지안, CUSUM 방법으로 유행을 탐지해 보자.

표 2-2-4. 베를린 A형간염 자료에서 유행을 탐지하기 위한 R 코드: CDC, 패링턴, 베이지안, CUSUM 알고리듬

```
# CDC 알고리듬을 이용한 유행의 감지
> ha.cdc <- algo.cdc(aggregate(ha), control = list(range = 209:290, b=3, w=5, alpha = 0.05))
# 유행감지의 범위는 2005~2006년(209주~290주)으로 설정. 과거 3개년의 자료를 이용(b=3), 자료분석의 넓이 5주(w=5), 유의수준 0.05로 control 매개변수를 설정
> plot(ha.cdc)                                                    # 결과 그래프

# 패링턴 알고리듬을 이용한 유행의 감지
> ha.farrington <- algo.farrington(aggregate(ha), control = list(range = 209:290, b=3, w=5, alpha = 0.05))
> plot(ha.farrington)

# 베이지안 알고리듬을 이용한 유행의 감지
> ha.bayes <- algo.bayes(aggregate(ha), control = list(range = 209:290, b=3, w=5, alpha = 0.05))
> plot(ha.bayes)
```

```
# CUSUM 알고리듬을 이용한 유행의 감지

> kh <- find.kh (ARLa = 500, ARLr = 7)
# spc 패키지 추가설치 필요. kh는 경계선을 결정하는 데 사용하며 이 이상 넘어가면 이상치로 감지. ARLa는 정상일 때의 평균 Run 길이, ARLr은 이상상태일 때의 평균 Run길이로 각각 조절가능.

> ha.cusum <- algo.cusum(aggregate(ha), control = list(k = kh$k, h = kh$h, m = "glm", trans = "rossi", range = 209:290))
# k는 CUSUM의 기준수치를 제공하기 위한 것이며, ARL기반의 값을 사용, h는 누적합이 이상치로 간주되는 임곗값으로 앞의 find.kh( ) 함수에서 출력된 값임. 방법은 일반선형모형(GLM)을 사용하고, 변환함수로는 Rossi 방법(관측값을 로그로 전환)을 사용.

> plot(ha.cusum)                                    # CUSUM결과의 출력
```

표 2-2-4은 CDC, 패링턴, 베이지안, CUSUM의 4가지 알고리듬으로 분석하는 코드(붉은색)이며 이 결과는 그림 2-2-6~9와 같다.

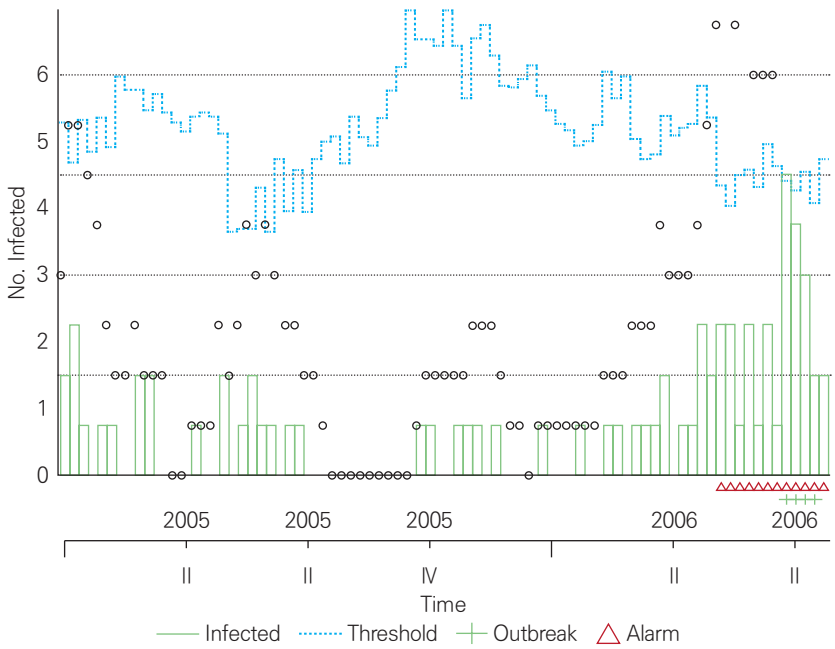


그림 2-2-6. CDC 알고리듬으로 감지한 A형 간염의 유행(붉은 삼각형)

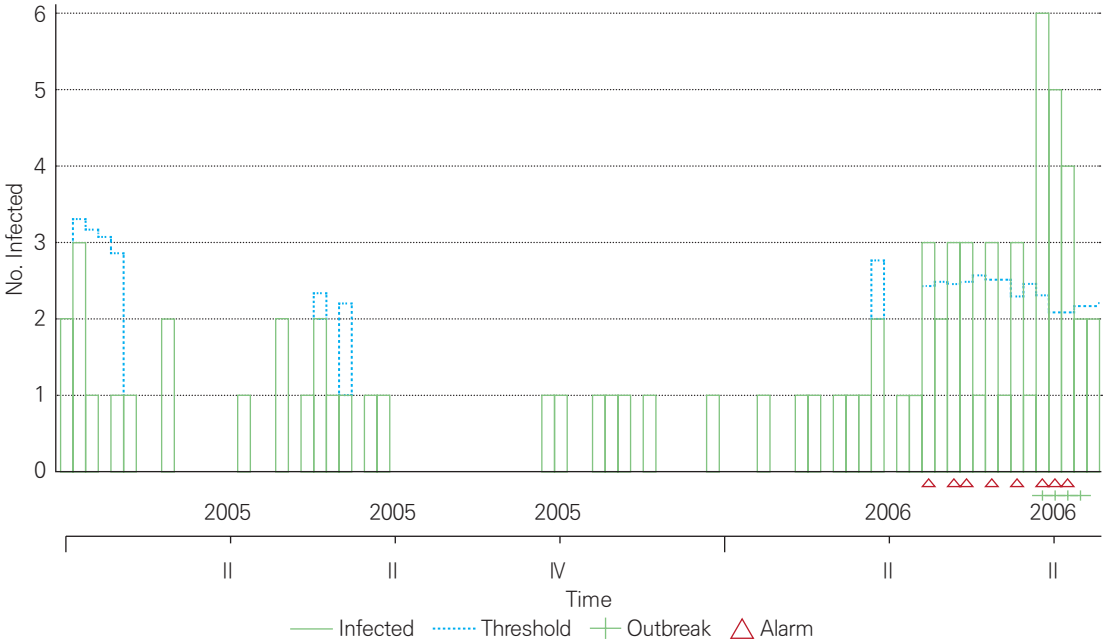


그림 2-2-7. 패링턴 알고리듬으로 감지한 A형 간염의 유행(붉은 삼각형)

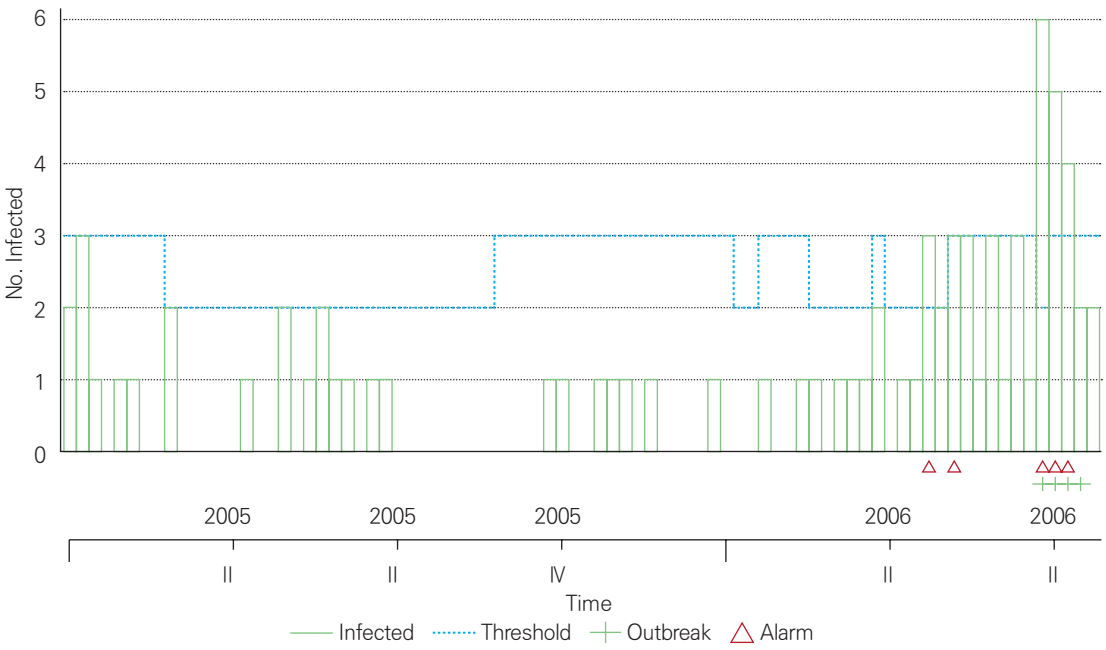


그림 2-2-8. 베이지안 알고리듬으로 감지한 A형 간염의 유행(붉은 삼각형)

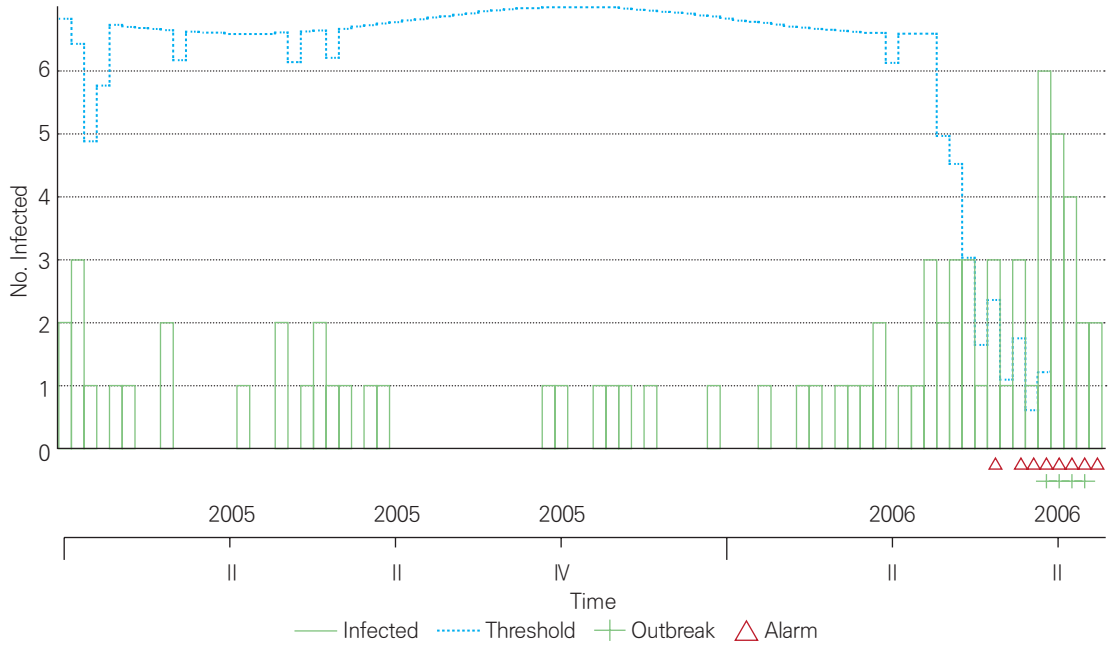


그림 2-2-9. CUSUM 알고리듬으로 감지한 A형 간염의 유행(붉은 삼각형)

그림 2-2-6은 CDC방법으로 유행을 감지한 것인데 실제 유행은 2006년 3/4분기(녹색+표)에 일어났지만, 이 보다 빠르게 2006년 2/4분기부터 유행에 대한 이상치를 감지하고 알람(붉은색 삼각형)을 보여주고 있다. 그림 2-2-7의 패링턴 방법도 유사하나 CDC방법보다 1주더 빠르게 이상치를 감지했다. 그림 2-2-7은 베이지안 방법인데 패링턴 방법과 거의 유사한 결과를 보여주고 있다. 그림 2-2-8은 CUSUM방법인데 CDC결과와 유사하다. 여기서 주의할 점은 이러한 이상치의 감지는 각 방법의 옵션에 따라서 달라진다는 것이다. 여기서도 표 2-2-4의 각 control 항목의 매개변수를 바꾸면 조금씩 다른 결과를 가져올 수 있다.

5 결론

R의 surveillance 패키지는 감시자료의 시계열 분석, 공간분석, 지도의 작성, 위험도 평가와 비교, 지역별 감염재생산수의 추정 등 매우 유익하고 다양한 분석을 할 수 있는 패키지이다. 역학조사관은 감시자료를 전문적으로 분석해야 하는 위치에 있으며, 이 패키지를 활용한다면 전문적 분석을 쉽게 할 수 있다.

그렇지만 이 패키지를 제대로 활용하기 위해서는 유행 감지를 위한 알고리듬을 이해해야 하며, R의 사용에 익숙해질 필요가 있다. 특히 surveillance 패키지를 이용해서 유행을 감지해 내기 위해서는 각 control 매개변수 옵션을 목적에 맞게 조절할 수 있어야 한다.

또한 R의 surveillance 패키지가 가지는 다양한 기능 중 여기서는 유행의 발견 기법만 보았기 때문에 다른 기능에 대해서도 관심을 갖고 익히면 감시자료 분석에 크게 도움이 될 것이다.

참고 문헌

1. CRAN [Internet]. surveillance: Temporal and Spatio-Temporal Modeling and Monitoring of Epidemic Phenomena [Cited 2025 Apr 16]. Available from: https://cran.r-project.org/web/packages/surveillance/index.html.
2. Meyer S, Held L, Höhle M. Spatio-Temporal Analysis of Epidemic Phenomena Using the R Package surveillance. J Stat Software. 2017;77(11):1-50.

Textbook For
Korea Field Epidemiology
Training Program





3장

유행역학조사 역량강화







제1절 유행곡선을 이용한 공통노출일과 평균잠복기 추정법
제2절 층화분석과 다변량분석을 이용한 연관성 지표의 보정
제3절 유행조사자료를 이용한 감염재생산수 추정법
제4절 수인성 및 식품매개감염병 유행 시 역추적 조사와 환경 평가의 활용

집필자 박지혁, 천병철

제1절 유행곡선을 이용한 공통노출일과 평균잠복기 추정법

유행곡선은 x축은 시간, y축은 시간단위별 새롭게 발생한 사례(new case)를 표시한 그래프이다. 유행곡선은 시간에 따른 유행의 기술방법으로 가장 많이 사용하고 필수처럼 여겨지는데 그것은 유행곡선을 통해서 알아낼 수 있는 정보가 매우 많기 때문이다. 그중의 하나가 공통노출일(common exposure day)을 추정하는 것과 평균 잠복기(mean incubation period)를 계산하는 것이다[참고: '공통노출(共通露出)'과 '공동노출(共同露出)'은 모두 쓰이고 있는데 '공통'이 같은 속성이나 특징을 공유하는 것에 초점을 맞춘 단어라면 '공동'은 여러 사람이 함께 일을 하거나 무엇인가를 함께 소유하는 상태를 의미한다. 공통된 오염원, 공통된 노출일자, 공통감염경로 등과 같이 공유하는 속성을 강조할 때는 공통을 쓰고, '공동으로 노출되었다'와 같이 함께하는 행위를 강조할 때는 공동을 사용하기도 한다. 여기서는 사례들이 공통적으로 노출된 시간라는 의미에서 공통노출일이라는 표현을 사용했다].

공통노출일이란 특정 감염원(source of infection)에 대한 노출이 발병 또는 감염 사례들에게 공통적으로 일어난 날짜 또는 시간을 의미한다. 여러 사람이 동시에 노출된 식중독이나 오염사고 등에서는 공통노출 시간을 파악하는 것은 매우 중요하다. 예를 들어서 한 학교에서 집단설사 사례가 신고되었다고 하자. 학생들은 매일 단체로 점심과 저녁 2끼를 함께 급식하고 있다. 유행규모 등을 볼 때 단체급식이 원인으로 의심된다고 할 때 언제 먹은 급식이 이 집단설사의 원인일지 조사하게 된다. 이때 음식에 대한 조사를 하루나 이틀 전까지만 해도 될지 아니면 1주일 식단을 다 조사해야 할지 아니면 그 이전의 것들도 조사해야 할지를 어떻게 결정할 수 있을까? 이미 사례들에서 분리된 병원체를 알고 있다면 그 병원체의 최장 잠복기를 이용해서 노출시기를 파악한 후 조사범위를 지정할 수 있지만, 역학조사 시 병원체가 아직 분리되기 전인 경우도 많고, 병원체가 분리되지 않는 경우도 있다. 이 경우에 유행곡선으로부터 공통노출일을 추산한다면 조사범위 설정이 쉬울 것이다.

또 예를 들어서 일별 식단에 대한 상대위험도를 조사했더니 여러 식단에서 통계적으로 유의한 연관성을 보이거나 같은 일자에 공통으로 섭취한 점심과 저녁이 모두 유의한 연관성을 보인다고 할 때 이 중의 어느 것이 이 유행의 진짜 원인인지를 공통노출일이나 공통노출시간을 계산함으로써 확인할 수 있게 된다. 또한 공통노출일을 알면 이 시기에 노출이 있었던 사람들을 모두 공통 감염원에 노출된 사람으로 보고 추적검사를 할 수도 있다.

1 단일공동노출의 유행곡선

유행곡선은 다양한 유행의 특성을 반영하기 때문에 유행곡선의 모양을 보고 전파의 특성을 파악할 수 있다. 유행곡선은 노출유형에 따라서 단일공동 노출형(point source exposure), 간헐 반복노출형(intermittent repeated exposure), 지속 노출형(continuous exposure), 확장전파형(propagative epidemic curve) 등으로 나눌 수 있다(표 3-1-1).

표 3-1-1. 노출유형에 따른 유행곡선의 특징

노출 유형	특징	곡선 모양	예시
단일공동노출 (Point Source Exposure)	• 공통된 감염원에 여러 사람이 동시에 한 번 노출됨	• 비대칭 종 모양(오른꼬리 분포)으로 처음에 급격한 증가 후 서서히 감소하는 모양	• 단체급식 시 오염된 음식으로 인한 식중독 • 작업장에서 화학 물질의 누출로 인한 집단 중독
간헐 반복 노출 (Intermittent Repeated Exposure)	• 동일한 감염원이 일정 기간 반복적으로 노출됨 • 환자 발생이 들쭉날쭉함	• 여러 개의 작은 피크가 나타남(노출 시 유행이 간헐적으로 반복)	• 간헐적으로 오염되는 식수 공급 • 간헐적으로 발생하는 공기 오염
지속 노출 (Continuous Exposure)	• 감염원에 장기간 지속적으로 노출됨 • 환자가 일정 기간 동안 계속 발생	• 완만하게 증가 후 고원형의 정체기(plateau) 형성, 감염원이 제거되면 감소	• 지속적으로 오염된 식수섭취로 인한 수인성 감염병(세균성이질 등) • 지속적 공장 배출 오염물질에 의한 호흡기 질환
확장전파형 (Propagative Epidemic)	• 사람 간 전파로 인해 유행이 확산됨 • 시간이 지남에 따라 환자가 점점 증가	• 여러 개의 점진적인 파동(waves)이 발생	• 홍역, 인플루엔자, COVID-19 등 호흡기 감염병

이 중에서 단일공동노출에 의한 유행(point source outbreak) 시 공통노출일을 추정할 수 있다. 다른 노출유형은 노출이 여러 번 일어난 것이기 때문에 하나의 공통노출일을 가지지 않는다. 단일공동노출이란 공통된 동일한 감염원에 여러 사람들이 동시에 한 번 노출된 경우를 의미하며 쉬운 사례가 학교나 회사와 같이 단체급식 시 특정 오염된 음식에 의한 식중독이나 집단설사 같은 경우가 해당된다.

단일공동노출의 전형적인 유행곡선은 그림 3-1-1과 같다. 실제 유행사례에서 흔히 보는 유형으로 사례 수가 급격히 증가하다가 정점 이후 서서히 감소하는 특징을 가진다. 이것은 단일공동노출로 이루어진 감염병의 유행 시 거의 예외 없이 나타나는데 그 이유는 대부분의 감염병의 잠복기는 로그정규분

포(log-normal distribution)를 하기 때문이다. 로그정규분포곡선은 급격히 증가 후 오른쪽으로 길게 꼬리를 가지는 오른꼬리 분포(right-skewed distribution)를 가진다.

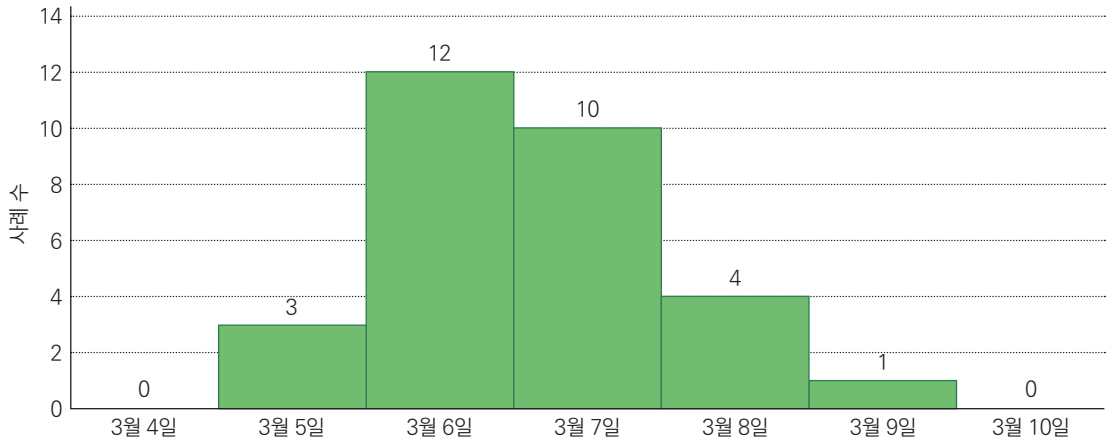


그림 3-1-1. 단일공동노출에 의한 집단설사의 유행곡선의 예

그림 3-1-1은 단일공동노출로 동시에 감염원에 노출된 사례이므로 3월 5일에 발병한 3명은 이 질환의 최단 잠복기에 해당하는 사람들이고 3월 9일에 발병한 사람은 최장 잠복기에 발병한 사람이다. 즉 이 유행곡선은 해당 감염병의 잠복기의 분포를 보여주고 있고, 잠복기는 로그정규분포를 하기 때문에 이와 같은 형태를 보여준다. 따라서 만약 x축의 시간에 로그를 취해서 x축을 로그 스케일로 바꾸면 이 분포는 좌우대칭 종모양의 정규분포로 바뀌게 된다. 이 특성을 이용하면, 비단 식품매개질환뿐 아니라 호흡기감염병, 오염으로 인한 환경성질환 등 한번에 여러 사람에게 동시 노출된 경우에는 모두 이 방법으로 공통노출일을 구할 수 있다.

2 유행곡선에서 공통노출일 계산원리

유행곡선을 이용해서 공통노출일을 계산하는 원리는 잠복기가 로그정규분포를 한다는 것을 이용한 것이다. 그림 3-1-2는 정규분포의 특성을 이용해서 공통노출일을 계산하는 원리를 표현한 것이다. 정규분포는 평균(μ)을 기준으로 좌우대칭 모양을 갖는다. 만약 평균에서 1 표준편차(σ)만큼 빼거나($\mu-\sigma$) 더하면($\mu+\sigma$) 정규분포의 특성에 따라서 이 선 밖으로는 각각 약 16%씩의 분포가 있게 된다. 즉 μ는 전체

환자의 50%가 발생하게 되고 이 시점을 m이라고 하고, $\mu-\sigma$는 전체 환자의 16%가 발생한 시점(a)이 되고, $\mu+\sigma$는 전체환자의 84%가 발생한 시점(b)이 된다. 이때 a와 b는 똑같이 μ에서부터 σ만큼 떨어진 거리므로 m-a=b-m가 된다(그림 3-1-2에서 $\alpha=\beta$).

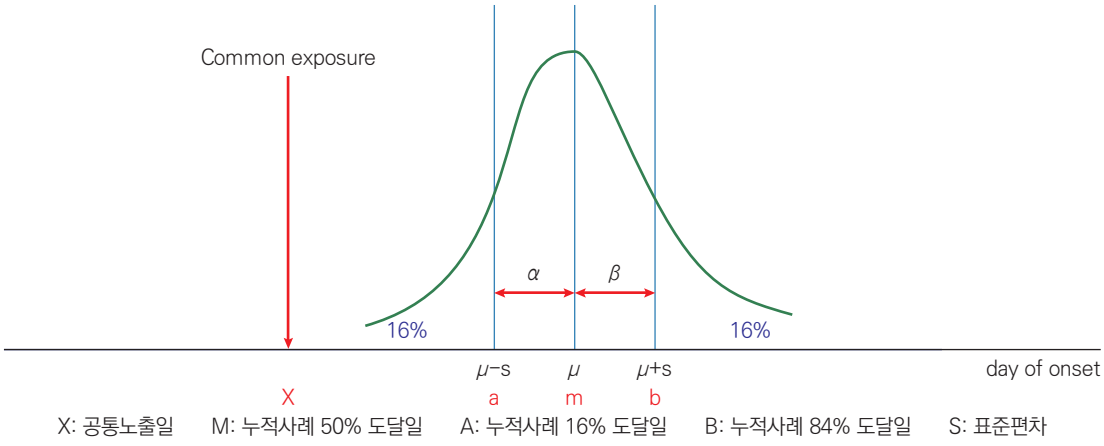


그림 3-1-2. 유행곡선에서 공통노출일 계산원리

이때 x를 사례들이 오염원에 동시에 노출된 공통노출일이라고 하자. 이것은 알려져 있지 않은 미지의 날짜이다. 이때 그림의 α는 $(m-x)-(a-x)$가 되고 β는 $(b-x)-(m-x)$가 된다. 이 둘은 같다. 그런데 실제 유행곡선은 정규분포가 아니라 로그정규분포이므로 각각의 값에 로그를 취해서 정규분포로 만들어서 계산을 해야 한다. 즉, 위의 내용은 식 1처럼 정리된다.

$$\text{식 1: } ln(m-x)-ln(a-x)=ln(b-x)-ln(m-x)$$

식 1을 x에 대해서 풀면 식 2가 된다.

$$\text{식 2: } x=(ab-m^2)/(a+b-2m)$$

즉 공통노출일을 계산하기 위해서는 전체 사례의 16%가 발생한 시점(a), 50%가 발생한 시점(m), 84%가 발생한 시점(b)만 알면 된다. (참고: 이것은 정규분포가 평균을 중심으로 좌우대칭이라는 것을 이용한 것이므로 꼭 16%, 84%만 기준으로 해야 하는 것은 아니다. 또 식 1에서 자연로그 ln 대신 상용로그 log를 취해도 식 2는 똑같이 얻어진다.)

3 공통노출일 계산 사례

그림 3-1-3은 집단급식후 발생한 설사환자의 유행곡선이다. 이 집단은 점심과 저녁 매 끼니를 단체로 급식하고 있었고 6월 7일 6명의 발병을 시작으로 총 141명이 발생하였다.

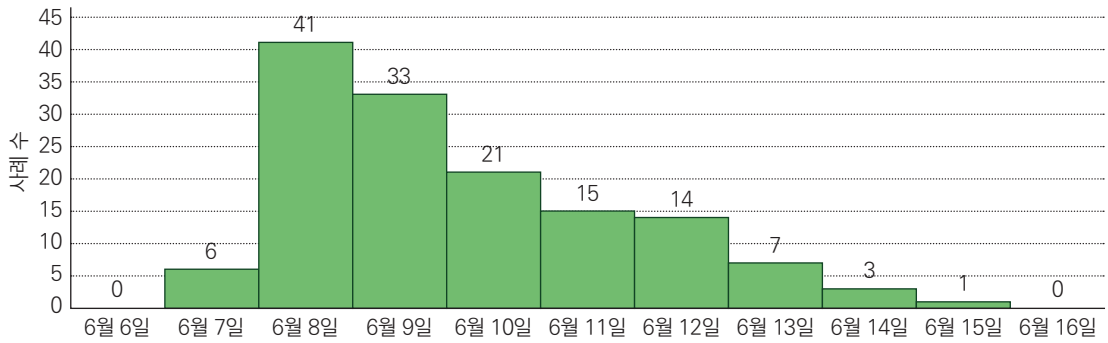


그림 3-1-3. 공통노출일을 계산하기 위한 유행곡선 사례

그림 3-1-3의 유행곡선을 이용해서 공통노출일을 추정하기 위해서는 식 2의 a, m, b의 값을 알아야 한다. 이것은 표 3-1-2를 만들어서 비례식으로 간단히 파악할 수 있다.

표 3-1-2. 집단설사 유행의 일별 사례수와 누적사례수, 누적사례분율

날짜	사례 수	누적사례수	누적사례분율
6월 6일	0	0	0.0
6월 7일	6	6	4.3
6월 8일	41	47	33.3
6월 9일	33	80	56.7
6월 10일	21	101	71.6
6월 11일	15	116	82.3
6월 12일	14	130	92.2
6월 13일	7	137	97.2
6월 14일	3	140	99.3
6월 15일	1	141	100.0
6월 16일	0	141	100.0

우선 a는 전체 사례의 16%가 발생하는 시점인데 6월 7일까지 4.3%, 6월 8일까지 33.3%가 발생했으므로 이는 6월 8일의 어느 시점이다. 마찬가지로 50%의 환자가 발생하는 시점은 6월 9일의 어느 시점이고 84%의 사례가 발생한 것은 6월 12일의 어느 시점이다. 이를 간단한 비례식으로 표시하면 아래와 같다.

a=6월 8일+(16-4.3)/(33.3-4.3)=6월 8.40일
m=6월 9일+(50-33.3)/(56.7-33.3)=6월 9.71일
b=6월 12일+(84-82.3)/(92.2-82.3)=6월 12.10일

식 2에다 위의 a, m, b값을 대입하면

$$x=(ab-m^2)/(a+b-2m)$$
$$=(8.40\times12.10-9.71^2)/(8.40+12.10-2\times9.71)$$
$$=6.81$$

즉, 6월 6.81일이 된다. 0.1일이 2.4시간이므로 6.81일은 6일 19.44시간이 되고 0.44시간을 분으로 환산하면 26.4분(=0.44×60)이므로 공통노출시각은 6월 6일 19시 26분으로 계산이 된다. 이것은 계산식이므로 분까지 계산이 되나 이것은 사례 1, 2명의 차이로도 많이 변하기 때문에 참고로만 하고 계산된 시간 전후의 노출을 함께 조사하는 것이 좋다. 즉, 이 유행의 원인이 된 것은 6월 6일 석식일 가능성이 아주 높다.

4 유행곡선으로 평균 잠복기 계산하기

유행곡선으로 평균 잠복기(mean incubation period)를 구하는 것은 간단한 일이 아니다. 만약 유행곡선이 정규분포를 하고 노출일을 이미 알고 있다면 노출일로부터 각 사례까지의 잠복기를 구하고 이것의 평균을 구하면 되지만, 아쉽게도 유행곡선은 정규분포가 아니라 로그정규분포를 한다. 만약 노출일을 알고 있다고 해도 기하평균을 구해야 한다. 평균 잠복기를 구하는 것이 쉽지 않기 때문에 WHO에서 나온 식품매개질환역학조사 매뉴얼을 보면 잠복기의 중위수(median)를 기술하도록 하고 있다. 중

위수는 전체 사례 중 50%가 발생한 시점이므로 현장에서도 쉽게 구할 수 있기 때문이다. 그러면 평균 잠복기는 어떻게 계산하는지를 앞의 사례를 가지고 알아보자.

표 3-1-3은 위의 공통노출일 계산 사례를 이용해서 평균 잠복기를 구하기 위한 방법이다. 환자가 발생한 6월 7일 이전부터 계산해서 기간을 일로 정리하고 이것의 로그값을 취한다(여기서는 자연로그를 사용하였다). 각 기간은 하루의 중간 지점을 기준으로 하였기 때문에 0.5일, 1.5일이 된다.

표 3-1-3. 평균 잠복기를 구하기 위한 누적사례분률과 기간에 대한 로그값

날짜	사례 수	누적사례수	누적사례분율	기간(day)	ln (day)
6월 6일	0	0	0.0	0.5	−0.69
6월 7일	6	6	4.3	1.5	0.41
6월 8일	41	47	33.3	2.5	0.92
6월 9일	33	80	56.7	3.5	1.25
6월 10일	21	101	71.6	4.5	1.50
6월 11일	15	116	82.3	5.5	1.70
6월 12일	14	130	92.2	6.5	1.87
6월 13일	7	137	97.2	7.5	2.01
6월 14일	3	140	99.3	8.5	2.14
6월 15일	1	141	100.0	9.5	2.25
6월 16일	0	141	100.0	10.5	2.35

이 유행분포의 평균은 m에 해당하는 기간(day)의 값이므로 다음과 같은 비례식으로 구할 수 있다. 6월 8.99일과 9.99일 동안 ln (day)가 0.92에서 1.25로 변하였으므로 이를 이용해서 m(앞서 6월 9.71일로 계산하였다)에 해당하는 ln (da)값을 찾으면 된다. 이를 식으로 표시하면

식 3: (9.99-8.99):(1.25-0.92)=(9.71-8.99):(x-0.92)

식 3을 x에 대해서 풀면 x=1.1576이 되고, 평균은 로그를 벗겨야 하므로 exp (1.1576)=3.182가 되어 평균은 약 3.2일이 된다.

이 유행분포의 표준편차는 얼마일까? 이것은 약 16%의 사람이 발생하는 시점(a)을 가지고 위와 같은 식으로 계산하면 된다.

식 4: (8.99-7.99):(0.92-0.42)=(8.40-7.99):(x-0.41)

식 4를 x에 대해서 풀면 x=0.615이 되고, 따라서 exp(0.615)=1.85가 되는데 이것이 표준편차에 해당하는 분산계수(dispersion factor)가 된다.

위에서 구한 평균과 표준편차를 이용해서 이 유행 잠복기의 95%신뢰구간을 구할 수 있다. 다만 여기서는 로그정규분포이므로 다음과 같은 기하평균과 기하표준편차의 공식을 이용한다.

$$95\%\ CI(하한, 상한)=(\mu/\sigma^{1.96}, \mu\times\sigma^{1.96})$$
$$=(3.182/1.85^{1.96}, 3.182/1.85^{1.96})$$
$$=(1.0, 10.6)$$

즉, 이 유행 잠복기의 95%신뢰구간은 1~10.6일이다. 이 유행은 세균성이질의 사례로 알려진 세균성이질의 잠복기에 잘 맞는다.

5 공통노출일과 평균잠복기 추산결과의 해석

위의 방법으로 공통노출일을 추정할 때 다음과 같은 주의가 필요하다.

① 이것은 단일공동노출이 잠복기의 분포를 나타내는 것을 이용한 것이므로 단일공동노출이 아닌 유행에서는 이 방법을 이용해서 공통노출일을 추산할 수 없다.

② 일차 발병자만으로 계산해야 한다. 만약 일차 발병자에 의해서 전파된 이차 발병자가 포함되면 오류를 야기한다.

③ 만약 알려지지 않은 이전의 유행에서 감염된 자가 포함되면 공동노출일이 엉뚱하게 계산될 수 있다. 그러나 통상의 유행곡선에서 첫 부분에 있는 사람이 같은 노출에 의해서 발생했는지 앞의 다른 노출에 의해서 발생했는지 알아보는 방법이 있다. 그것은 이 사람을 제거했을 때와 포함했을 때의 잠복기 분포를 비교해 보면 된다. 같은 유행인 경우는 세미로그 그래프(semi-logarithmic graph)를 그리면 누적분포가 직선을 이룬다(로그 정규분포이기 때문에). 그러나 만약 다른 유행에서 감염된 사람이 포함되면 이 직선에서 벗어나게 된다. 직선에서 벗어난 사람을 제거하고 계산하면 실제 유행과 잘 맞는 결과를 얻을 수 있다.

④ 이 방법을 이용해서 공통노출일을 계산하기 전에 잠복기가 로그정규분포를 하는지 점검하는 것

이 좋다. 만약 벗어나는 이상점(outlier)이 있다면 이 사례는 같은 공통감염원이 아닌 다른 감염원으로부터 전파되었을 수 있으므로 이를 제거하고 계산하는 것이 좋다. 로그정규분포를 하는지를 점검하는 것은 세미로그 그래프를 그려서 눈으로 확인하는 방법도 있고, 통계분석프로그램에서는 로그로 변환한 값들에 대해서 샤피로-윌크 검정(Shapiro-Wilk test), 콜모고로프-스미르노프 검정(Kolmogorov-Smirnov test)과 같은 다양한 정규성 검정 프로그램을 이용하면 된다.

⑤ 만약 계산결과가 실제 상황과 잘 맞지 않는다면 다음을 점검해 보자

- 사례정의가 객관적이지 않은 경우: 사례정의가 분명해서 오분류가 없어야 하는데, 만약 사례정의가 불분명해서 사례가 아닌 사람이 섞이거나 다양한 사례가 포함되면 잘못된 결과를 얻을 수 있다.
- 유행이 현재 계속되고 있는 상황(즉, 환자가 계속 발생하고 있는 상황)이라면 이 방법을 사용하기 어렵다.
- 유행곡선의 시간 단위: 시간단위를 잘 사용해야 한다. 통상 일(day)을 많이 쓰지만 잠복기가 짧은 경우는 일보다 시간(hour) 단위를 쓰게 된다. 시간 단위를 바꾸면 분포 모양이 바뀌고 로그정규분포를 잘 나타내는 시간단위를 먼저 찾은 후 이 단위를 이용해서 공통노출시간을 찾아야 한다.
- 발병일에 기초해서 유행곡선을 그려야 하는데, 발병일이 없이 진단일이나 신고일로 유행곡선을 그리거나 이 둘이 함께 섞여 있는 유행곡선으로는 공통노출일을 구할 수 없다. 발병일로 그려진 유행곡선을 이용해야 하며, 발병일이 없는 무증상감염자는 제외하고 그린 후 계산해야 한다.
- 발병일은 대부분 기억에 의존하기 때문에 발병일 이용 시 이에 대한 한계는 존재하나 통상 이로 인한 바이어스는 전체적인 추산에 크게 영향을 주지 않는다. 그렇지만 사례의 누락이나 다른 유행사례가 포함된 경우는 추정일에 적지 않게 영향을 준다.

이 방법은 실제 역학조사 상황에서 유용하게 사용할 수 있음에도 잘 알려지지 않아서 사용이 안 되고 있는 것 중의 하나이다. 잘 익혀서 실제 자료에 활용해 보자.

참고 문헌

1. 고응린. 계량의학통론. 서울: 신광출판사; 1989.
2. 김정순. 역학원론. 서울: 신광출판사; 2000.

3. 천병철. 유행곡선을 이용한 공통노출일의 추정. 계량역학 강의록. 고려대학교 보건대학원. 2024

제2절 층화분석과 다변량분석을 이용한 연관성 지표의 보정

1 연관성 지표와 교란인자

역학에서 사용하는 대표적인 연관성 지표는 상대위험도(relative risk, RR), 오즈비(Odds ratio, OR) 그리고 위험비(hazard ratio, HR)가 있는데 HR은 사실상 RR과 같은 의미이고, 오즈비는 제한적 상황에서 (발병률이 10% 미만) RR처럼 해석을 할 수 있다. 이들은 모두 특정 노출과 질병(또는 사망)과의 연관성을 측정하는 지표들이다. 그런데 관찰역학연구에서 연관성을 추정할 때는 많은 제3의 요인들이 연관성 추정에 영향을 미친다.

예를 들어서 남녀 간에 폐렴 발생에 차이가 있는지를 파악하기 위한 사례-대조군연구를 수행해서 표 3-2-1과 같은 결과를 얻었다고 하자. 여기서 폐렴사례와 대조군 각각 177명씩 조사했더니 폐렴사례군에서 남자의 비율이 더 많았고 이때 오즈비는 2.08이었다. 이 결과만 보면 남자에서 폐렴에 걸릴 위험이 2배나 더 높아 보인다. 표 3-2-1의 결과를 연령 65세 이상과 그 이하로 층화해서 다시 분석해 보면 표 3-2-2와 같다.

표 3-2-1. 폐렴과 성별의 연관성에 대한 사례-대조군연구 결과

	폐렴사례군	대조군
남	102	70
여	75	107
합계	177	177

OR=2.08

표 3-2-2에서 보는 것처럼 표 3-2-1의 자료를 연령으로 층화해서 폐렴과 남녀의 연관성을 파악하면 65세 이상군이나 65세 미만군에서 모두 OR=1.00으로 남녀 간에 차이가 없음을 알 수 있다. 즉, 성별은 폐렴 발생의 위험요인이 아닌 데도 표 3-2-1에서는 남자에서 2.1배의 위험이 있는 것으로 추산된

것이다. 이때 연령으로 층화한 오즈비는 연령의 효과를 보정한 오즈비라고 할 수 있다.

표 3-2-2. 연령을 65세 이상과 미만으로 층화하였을 때 폐렴과 성별의 연관성

	연령 65세 이상			연령 65세 미만	
	폐렴사례군	대조군		폐렴사례군	대조군
남	72	16	남	30	54
여	18	4	여	57	103
합계	90	20	합계	87	157

65세 이상 OR=1.00, 65세 미만 OR=1.00

표 3-2-1과 표 3-2-2의 자료를 남녀 구분 없이 연령에 따라서만 나누어 보면 표 3-2-3과 같이 된다. 표 3-2-3에서 보는 것과 같이 폐렴과 연령의 연관성에서는 연령이 65세 이상인 군의 오즈비가 8.12로 고연령층에서 폐렴의 위험이 매우 높음을 알 수 있다. 즉, 65세 이상의 고령은 폐렴 발생의 위험요인인 것이다.

표 3-2-3. 폐렴과 연령의 연관성

	폐렴사례군	대조군
연령 65세 이상	90	20
연령 65세 미만	87	157
합계	177	177

OR=8.12

표 3-2-1, 2, 3은 같은 자료를 성별, 연령, 또는 연령으로 층화한 성별로 나누어 본 것으로 결론적으로 폐렴은 연령과 연관성이 있고 성별은 연관이 없는 것을 알 수 있다. 그런데도 표 3-2-1에서는 성별이 연관성이 있는 것처럼 나왔는데 이는 표 3-2-2에서 보는 것처럼 남자에서 65세 이상이 많이 포함되고 상대적으로 여자에서 65세 미만이 많이 포함된 데서 오는 착시로 이를 교란효과(confounding effet)라고 하며, 이렇게 교란작용을 일으키는 것을 교란인자(confounding factor, confounder) 또는 교란변수(confounding variable)라고 한다. 따라서 폐렴과 성별의 연관성 오즈비 2.0은 착시이며 이에 대한 연령의 효과를 보정하지 않으면 연관성의 참값을 알 수 없다. 이렇게 연관성에 영향을 주는 요인을 제거하는 것을 보정(adjustment)이라고 하며 최종 결과로 보정된 오즈비(adjusted OR)를 제시해야 한다. 상대적으로 보정을 하지 않은 오즈비는 조오즈비(crude OR) 또는 비보정 오즈비(non-ad-

justed OR)라고 한다.

교란인자는 이렇게 노출과 결과(질병)에 포함되지 않은 제3의 요인으로 다음의 3가지 조건을 만족시킨다. 또는 반대로 노출과 결과의 연관성에서 이 3가지 조건을 만족시키는 제3의 요인을 교란인자라고 정의한다.

- 노출과 독립적으로 교란인자 자체가 질병의 위험요인이다.
- 교란인자는 노출과 연관성을 가지고 있다.
- 교란인자는 노출과 결과 사이에 있는 중간매개인자(intermediate factor)가 아니다.

이러한 교란인자의 조건을 그림 3-2-1과 같이 표시할 수 있다.

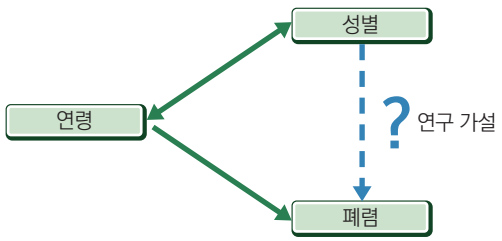


그림 3-2-1. 성별과 폐렴의 연관성에 대한 연구에서 연령의 교란효과

여기서 성별이 원래 연구하고자 했던 노출이고 폐렴이 결과이다. 연령은 제3의 요인으로서 이 연관성에 영향을 주는 교란인자이다.

2 연관성 지표와 효과수정인자

앞에서 본 교란변수와 유사한 존재로 연관성 측정에 영향을 주지만 태생적으로 다른 것이 있는데 이것이 효과변경인자 또는 효과수정인자(effect modifier)이다. 결론부터 이야기하면, 교란변수는 층화분석이나 다변량분석을 통해서 반드시 보정해 주어야 참연관성 지표값이 나오지만, 효과변경인자는 다변량분석에 포함해서 보정해서는 안 된다. 교란변수는 연관성 측정에 착시를 일으키고 왜곡하기 때문에 반드시 제거나 보정이 필요한 반면, 효과변경인자는 연구자에게 도움을 주는 인자로 이 인자의 하위군

별로 연관성이 어떻게 달라지는지 참값을 알려주는 존재라고 할 수 있다(표 3-2-4).

표 3-2-4. 교란인자와 효과수정인자의 차이

특징	교란인자(Confounder)	효과수정인자(Effect Modifier)
정의	연구 대상인 노출과 결과 모두에 영향을 미치는 외부 변수	노출과 결과 사이의 연관성 강도 또는 방향을 변화시키는 외부 변수
역할	노출과 결과 사이의 관계를 왜곡하여 실제 관계를 숨기거나 거짓 관계를 만들어 냄	노출과 결과 사이의 관계가 서로 다른 하위 그룹에서 어떻게 달라지는지 설명
목적	연구 결과의 타당성을 떨어뜨림	연구 결과의 일반화 가능성을 높이고, 특정 하위 그룹에 대한 추가적인 정보를 제공
통계적 분석	층화 분석, 다변량 분석 등을 통해 통제 또는 보정	층화 분석을 통해 하위 그룹별로 효과의 차이를 분석
예시	흡연과 폐암 연구에서 '나이'는 흡연과 폐암 모두에 영향을 미치므로 교란변수	흡연과 폐암 연구에서 '성별'은 흡연이 폐암에 미치는 영향의 강도를 변화시키므로 효과수정변수
연구자의 역할	교란인자를 식별하고 통제하여 편향을 최소화	효과수정인자를 식별하고 하위 그룹별 효과의 차이를 분석하여 연구 결과를 해석

교란인자와 효과수정인자의 구분하기 위한 정의를 그림으로 설명하면 그림 3-2-2와 같다.

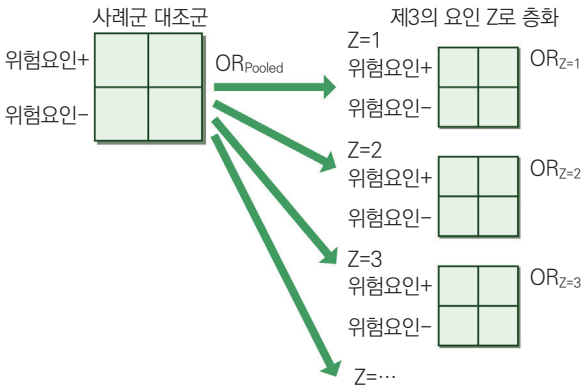


그림 3-2-2. 제3의 요인으로 층화한 오즈비를 이용한 교란인자와 효과수정인자의 정의와 구분

그림 3-2-2는 사례-대조군연구에서 OR을 추정할 때 제3의 인자 Z에 대해서 층화한 것을 나타낸 것으로 층화하기 전의 전체 자료의 OR을 OR_{pooled}로 표시하고, 각 층화된 표에서 계산된 OR을 $OR_{z=1}$, $OR_{z=2}$ 등으로 표시한 것이다. 예를 들어서 연령으로 층화했을 때 65세 미만과 65세 이상의 각 군에서 계산된 OR이 여기에 해당된다. 이때 교란인자와 효과수정인자는 각각 아래와 같이 정의된다.

- $OR_{pooled}=(OR_{z=1}=OR_{z=2}=OR_{z=3}=\cdots)$ → Z는 교란인자나 효과수정인자가 아님
- $OR_{pooled}\neq(OR_{z=1}=OR_{z=2}=OR_{z=3}=\cdots)$ → Z는 교란인자임
- $OR_{pooled}\neq(OR_{z=1}\neq OR_{z=2}\neq OR_{z=3}\neq\cdots)$ → Z는 효과수정인자임

만약 층화하기 전 OR_{pooled}이 각 층화된 표에서의 $OR_{z=1}$ 등과 같다면 Z는 교란인자나 효과수정인자가 아니기 때문에 그대로 OR_{pooled}를 제시하면 된다. 그런데 만약 Z로 층화된 각 오즈비는 서로 같은데 (즉, $OR_{z=1}=OR_{z=2}=OR_{z=3}=\cdots$) 이들이 OR_{pooled}과 다르면 이때 Z는 교란인자로 반드시 보정된 OR을 제시해 주어야 한다. 그런데 만약 층화된 OR이 다 다르고(즉, $OR_{z=1}\neq OR_{z=2}\neq OR_{z=3}\neq\cdots$) 이들의 OR들이 OR_{pooled}과 다르다면 이때의 Z는 효과수정인자로서 층화분석된 OR들을 제시하고 전체 보정된 OR을 제시해서는 안 된다. 최근 통계패키지가 쉽게 사용되면서 교란변수와 효과수정인자의 구분 없이 제3의 요인을 모두 다변량분석에 포함해서 분석하는 경향이 있는데 이는 제대로된 연관성 지표 보정방법이 아니다. 다변량 분석에 포함시키는 것은 기본적으로 교란인자이고, 효과수정인자는 층화분석결과를 제시해 주어야 한다.

3 교란인자를 보정한 연관성지표의 계산

1. 연구설계 시 보정법

교란인자를 보정하기 위해서 연구설계 단계에서부터 연구대상자를 제한하거나 (연구대상 제한법, restriction), 연구대상자들을 교란인자에 대해서 짝짓기(matching)하는 방법 등이 있다. 표 3-2-1의 사례에서 연령이 교란인자라면 처음부터 65세 미만 또는 65세 이상만으로 사례군과 대조군을 구하는 방법이 제한법이다. 짝짓기는 사례군과 대조군을 뽑을 때 연령을 짝을 지어서 사례군과 대조군의 연령비율을 맞추는 방법이다. 여기에는 개별짝짓기와 빈도짝짓기는 방법이 있는데 짝짓기는 여러 변수에 대해서 짝지을 수가 없고, 분석에서 짝지은 자료의 분석방법만 써야 하는 등의 제한점이 있다.

일반적인 역학조사의 상황에서 상당한 시간을 가지고 사례-대조군을 수행하는 과정이 아니라면 현장에서 연구설계 시 보정하는 방법을 쓰는 것은 쉽지 않을 수 있다. 연구대상의 제한이나 짝짓기도 교란변수에 해당하는 변수만 써야 하며 효과수정인자에 대해서는 이러한 방법을 써서는 안 된다. 즉 확실한 교란변수라고 미리 알고 있는 경우에 해당하는 방법이다.

연구설계 시 교란변수가 보정이 되었다면, 자료분석 시 따로 보정하지 않아도 되고 그대로 OR이나 RR을 제시하면 되는데 짝지은 경우는 반드시 짝비교에 해당하는 통계분석방법으로 계산해야 한다.

2. 층화분석과 맨텔-핸젤 보정법

대부분의 현장역학에서는 연구설계 시 교란변수를 고려하여 접근하는 것이 쉽지 않고, 관찰된 자료를 수집 후 연관성 계산 시 보정을 하게 된다. 이때 보정하는 방법은 층화분석(stratification)과 다변량분석(multivariate analysis)이 있다.

층화분석은 직관적이고 교란변수의 수가 적을 때 사용한다. 교란변수의 수가 많으면 층화분석으로 보정하기가 어렵다.

예를 들어서 다음의 살모넬라 유행역학조사 자료를 보자. 여기서 가장 의심이 되는 음식 두 가지는 돼지고기와 달걀이었다. 표 3-2-5에서 돼지고기의 상대위험도(RR)는 1.50(=0.60/0.40)이고 달걀의 상대위험도는 1.97(=0.74/0.38)이며 둘다 통계적으로 유의하였다($p<0.05$). 그러면 둘 다 원인일까? 표 3-2-6과 같이 달걀섭취를 이용해서 층화분석을 해보자.

표 3-2-5. 살모넬라 식중독 역학조사 결과

음식	섭취			비섭취		
	사례수	전체	발병률	사례수	전체	발병률
돼지고기	180	300	0.60	20	50	0.40
달걀	170	230	0.74	30	80	0.38

표 3-2-6을 보면 원인은 달걀이라는 것을 알 수 있다. 왜냐하면 달걀섭취군은 돼지고기를 섭취하든 안 하든 발병률이 84%, 75%로 매우 높지만, 달걀비섭취군에서는 돼지고기 섭취한 경우 발병률이 15%, 돼지고기 섭취 안 한 군이 30%으로 낮아지고 오히려 돼지고기 섭취군에서 발병률이 더 낮기 때문이다. 이때 달걀섭취군에서의 돼지고기 상대위험도는 1.12(=0.84/0.75)이고 달걀비섭취군에서의 돼지고기 상대위험도는 0.5(=0.15/0.3)인데 이 상대위험도가 달걀섭취를 보정한 각 층의 돼지고기의 상대위험도가 된다. 즉 표 3-2-5에서 돼지고기의 상대위험도가 1.5로 유의하게 나온 것은 달걀섭취의 교란효과로 인한 착시였다.

표 3-2-6. 살모넬라 식중독에서 달걀섭취에 따른 층화분석 결과

구분	달걀섭취			달걀비섭취		
	사례 수	전체	발병률	사례 수	전체	발병률
돼지고기섭취	160	190	0.84	17	110	0.15
돼지고기 섭취 안 함	30	40	0.75	3	10	0.30

이와 같이 층화분석이라도 하지 않으면 달걀이 원인인지 돼지고기가 원인인지 알 수가 없게 된다. 이것이 현재 많은 식품매개질환 역학조사에서 원인을 특정하지 못하는 중요한 이유 중의 하나이다. 이와 같은 층화분석을 시도한 역학조사는 현재 역학조사서 중에서 손에 꼽을 정도로 드문데 정말 큰 문제가 아닐 수 없다.

예를 들어서 2019년 부산에서 유행한 A형간염의 다음 조사결과를 보자(교재 제3권 제2편 역학조사 사례 1-19-1 참조). 표 3-2-7은 식단별 상대위험도를 구한 것으로 조개젓의 상대위험도가 96.2로 매우 높지만 갓김치, 새우젓깔 등의 상대위험도도 2~3 정도이며 통계적으로 유의한 것을 볼 수 있다. 통상 역학조사에서 2, 3 정도의 상대위험도면 원인 음식일 가능성이 크다. 이 때문에 이 역학조사팀에서는 이들 음식에 대해서 조개젓으로 층화한 후 다시 상대위험도를 계산하였다(표 3-2-8). 표 3-2-8에서 보는 것처럼 조개젓을 먹은 군과 먹지 않은 군에서 다시 해당 음식에 대해서 상대위험도를 구했을 때, 상대위험도가 낮아지고 통계적 유의성이 사라진 것을 볼 수 있다. 이때 표 3-2-8의 각 음식의 달라진 상대위험도는 조개젓 섭취를 보정한 각 층의 상대위험도로 처음에 유의하게 나온 것은 조개젓의 교란효과 때문이라는 것을 알 수 있다. 즉, 이들은 이 식중독의 원인음식이 아님을 이 층화분석만으로도 보여줄 수 있는 것이다.

표 3-2-7. 2019년 부산 A형 간염 역학조사의 식단별 상대위험도

식단	섭취			비섭취			상대위험도(95% 신뢰구간)	
	대상자 수 (명)	사례 수 (명)	발병률(%)	대상자 수 (명)	사례 수 (명)	발병률(%)	상대위험도	CI: 하한~상한
오겹생대패	1,256	98	7.8	109	7	6.4	1.22	0.58~2.55
목살생대패	821	73	8.9	511	32	6.3	1.42	0.95~2.12
미전지생대패	636	62	9.7	669	40	6.0	1.63	1.11~2.39
파김치	1,204	100	8.3	146	5	3.4	2.43	1.00~5.86
갓김치	1,153	96	8.3	187	5	2.7	3.11	1.28~7.55
쌈장	1,203	99	8.2	158	5	3.2	2.60	1.08~6.29

식단	섭취			비섭취			상대위험도(95% 신뢰구간)	
	대상자 수 (명)	사례 수 (명)	발병률(%)	대상자 수 (명)	사례 수 (명)	발병률(%)	상대위험도	CI: 하한~상한
새우젓갈	587	64	10.9	677	36	5.3	2.05	1.38~3.04
조개젓갈	752	130	17.3	556	1	0.2	96.12	13.48~685.40
갈치속젓	611	68	11.1	652	33	5.1	2.20	1.47~3.28
참기름장	978	74	7.6	339	27	8.0	0.95	0.62~1.45
장아찌	935	71	7.6	346	29	8.4	0.91	0.60~1.37
상추	1,044	77	7.4	306	27	8.8	0.84	0.55~1.27
깻잎	1,038	80	7.7	308	24	7.8	0.99	0.64~1.53
고추	955	74	7.7	397	28	7.1	1.10	0.72~1.67
마늘	1,057	76	7.2	292	26	8.9	0.81	0.53~1.24
공깃밥	913	61	6.7	425	42	9.9	0.68	0.46~0.98
김치찌개	824	72	8.7	489	27	5.5	1.58	1.03~2.43
된장찌개	357	21	5.9	948	79	8.3	0.71	0.44~1.12
비빔면	90	6	6.7	1,247	98	7.9	0.85	0.38~1.88

표 3-2-8. 2019년 부산의 A형간염 역학조사에서 조개젓 섭취여부에 따른 음식의 상대위험도

조개젓 섭취 여부	식단	섭취			비섭취			상대위험도(95% 신뢰구간)	
		대상자 수(명)	사례 수 (명)	발병률 (%)	대상자 수(명)	사례 수 (명)	발병률 (%)	상대위험도	CI: 하한~상한
섭취	미전지생대패	376	60	16	314	40	12.7	1.25	0.86~1.82
	파김치	667	99	14.8	43	4	9.3	1.60	0.62~4.13
	갓김치	648	95	14.7	56	4	7.1	2.05	0.78~5.37
	쌈장	662	97	14.7	55	5	9.1	1.61	0.69~3.79
	새우젓갈	489	64	13.1	200	35	17.5	0.75	0.51~1.09
	갈치속젓	504	68	13.5	184	32	17.4	0.78	0.52~1.14
	김치찌개	455	71	15.6	240	27	11.3	1.39	0.92~2.10
비섭취	미전지생대패	207	1.5	0.7	315	0.5	0.2	4.57	0.19~111.53
	파김치	440	0.5	0.1	97	1.5	1.5	0.07	0.00~1.79
	갓김치	413	0.5	0.1	125	1.5	1.2	0.10	0.00~2.46
	쌈장	449	1.5	0.3	100	0.5	0.5	0.67	0.03~16.28
	새우젓갈	77	0.5	0.6	464	1.5	0.3	2.01	0.08~48.86
	갈치속젓	85	0.5	0.6	456	1.5	0.3	1.79	0.07~43.53
	김치찌개	319	1.5	0.5	215	0.5	0.2	2.02	0.08~49.40

층화분석을 하였을 때, 이렇게 교란변수가 있을 때와 없을 때의 오즈비나 상대위험도를 제시할 수도 있지만 이들을 통괄해서 보정한 오즈비나 상대위험도를 제시할 수도 있다. 이때 사용하는 것이 맨텔-핸젤 보정법(Mantel-Haenszel method)이다. 맨텔-핸젤 보정법에서 구한 보정된 OR을 OR_{MH}이라고 하면 다음과 같은 식에 의해서 구할 수 있다.

$$\text{식 1: } OR_{MH}=\frac{\sum_i \frac{a_i d_i}{N_i}}{\sum_i \frac{b_i c_i}{N_i}}$$

예를 들어서 표 3-2-1과 표 3-2-2의 자료를 이용해서 연령을 보정한 성별의 맨텔-헨젤 오즈비를 구하면 아래와 같다.

$$\text{식 2: } OR_{MH}=\frac{\frac{72\times 4}{110}+\frac{30\times 103}{244}}{\frac{16\times 18}{110}+\frac{54\times 57}{244}}=\frac{15.28}{15.23}=1.003$$

즉, 표 3-2-1에서 폐렴에 대한 성별의 오즈비는 2.0으로 남성에서 더 발생위험이 높은 것처럼 보였으나 연령별로 층화한 후 만텔-핸젤방법으로 연령을 보정한 오즈비는 1.00으로 성별간 폐렴위험의 차이가 없는 것을 알 수 있다. 즉, 처음 OR=2는 연령이라는 교란변수가 왜곡시킨 값이다. 이와 같이 보정없이 오즈비나 상대위험도를 제시하는 것은 왜곡된 값이 대부분이므로 교란변수를 보정한 값을 제시해야 한다.

3. 다변량분석법

다변량분석은 여러 개의 교란변수가 있을 때 이들 교란변수들을 한 번에 보정할 수 있는 방법이며 층화분석 대비 적은 수의 표본으로도 가능하며 분석 목적에 따라 다양한 모델을 만들 수 있기 때문에 가장 많이 사용하는 보정 방법 중의 하나이다. 이미 다양한 목적의 다변량분석법이 개발되어 있는데 예를 들어서 여러 변수를 보정한 오즈비를 얻고자 한다면 로지스틱 회귀분석을 통해서 얻을 수 있고, 마찬가지로 여러 변수를 보정한 상대위험도는 포아송 회귀분석을 통해서 얻을 수 있다. 또한 코호트연구와 같은 경시적 연구에서 많은 변수가 보정된 위험비(Hazard ratio, HR)를 얻고자 하면 콕스 회귀분석을 통해

서 얻을 수 있다. 이들은 매우 좋은 모델들이고 얻을 수 있는 결과도 풍부하지만 전문 통계프로그램을 이용해야 하는 단점이 있다. 그러나 최근에는 R, Python과 같은 무료 프로그램이 잘 발달되어 있고 이들의 사용법도 많이 보급되어 있어서 고가의 전문 프로그램이 없어서 분석을 못한다는 변명은 이제 더 이상 통하기 어렵다(이들 다변량 분석방법의 특징과 R을 이용한 분석방법은 제2권 2장을 참조하자).

다음은 2011년 서울의 한 학원에서 집단으로 발생한 클로스트리디움 퍼프린젠스 감염증역학조사서의 결과의 일부이다(제3권 제2편 사례 1-11-02 참조). 표 3-2-9는 3월 10, 11, 12일 저녁 식단별 오즈비를 제시하여 주고 있는데, 붉은색으로 표시한 것들이 통계적으로 유의하게 나온 음식들이다. 그런데 3월 10일과 11일 식단에서도 유의한 음식들이 많지만, 특히 3월 12일 저녁의 경우 돼지안심장조림의 오즈비는 35.5, 닭볶음탕의 경우는 40.6, 백미밥 29.8로 매우 높다. 이 역학조사에서는 이들 음식을 가지고 다변량 로지스틱분석을 해서 보정된 오즈비를 계산하였다(표 3-2-10). 표 3-2-10의 다변량분석 결과를 보면 닭볶음탕만 유의하고 다른 것들은 오즈비가 매우 낮아지고 통계적 유의성이 사라진 것을 볼 수 있다. 즉, 보정하지 않았을 때는 여러 음식들이 오염된 것처럼 보였으나 실제는 닭볶음탕 하나만 유의하고 나머지 음식들이 유의하게 보였던 것은 닭볶음탕의 교란효과 때문인 것이다 따라서 이러한 교란효과의 보정 없이 원인음식에 대한 오즈비나 상대위험도만 제시하는 경우, 그 결과는 오류일 가능성이 크다(다만 이 역학조사는 몇 개의 한계가 있는데 이에 대해서 제3권 제2편 사례 1-11-02 의견부분을 참조하자).

표 3-2-9. 한 학원에서 발생한 클로스트리디움 감염증에서 식단별 오즈비

식단 음식물		답변자 수	환례 수	환례		비환례		상대위험비 (오즈비)	95% CI		P
날짜	음식이름			섭취자	비섭취자	섭취자	비섭취자		Lower	Upper	
3/10 저녁	북엇국	289	110	84	26	113	66	1.887	1.106	3.220	0.027
	소불고기볶음	296	112	107	5	148	36	5.205	1.977	13.703	0.001
	미트볼볶음	299	113	109	4	154	32	5.662	1.946	16.475	0.001
	골뱅이무침	296	111	102	9	138	47	3.860	1.809	8.234	0.000
	부추장떡	295	112	97	15	140	43	1.986	1.045	3.775	0.049
	파래무침	294	111	83	28	101	82	2.407	1.434	4.040	0.001
	배추김치	297	111	92	19	120	66	2.663	1.494	4.747	0.001
	백미밥	297	112	110	2	153	32	11.503	2.700	49.011	0.000
3/11 저녁	김치찌개	291	107	87	20	136	48	1.535	0.854	2.761	0.196
	치즈돈까스	295	108	104	4	154	33	5.571	1.917	16.195	0.001
	골뱅이소면	296	109	101	8	148	39	3.327	1.492	7.416	0.004
	과일샐러드	295	109	92	17	139	47	1.830	0.990	3.382	0.072

식단 음식물		답변자 수	환례 수	환례		비환례		상대위험비 (오즈비)	95% CI		P
날짜	음식이름			섭취자	비섭취자	섭취자	비섭취자		Lower	Upper	
3/11 저녁	비엔나칠리볶음	295	109	106	3	152	34	7.904	2.366	26.404	0.000
	무생채	293	108	79	29	111	74	1.816	1.083	3.047	0.032
	깍두기	295	109	90	19	120	66	2.605	1.460	4.648	0.002
	백미밥	296	109	107	2	155	32	11.045	2.592	47.070	0.000
3/12 저녁	오뎅국	297	116	101	15	121	60	3.339	1.788	6.234	0.000
	돼지안심장조림	**301**	**115**	**114**	**1**	**142**	**44**	**35.324**	**4.793**	**260.327**	**0.000**
	고등어조림	301	115	97	18	128	58	2.442	1.352	4.409	0.004
	방울토마토	300	115	78	37	106	79	1.571	0.965	2.559	0.089
	닭볶음탕	**302**	**117**	**116**	**1**	**137**	**48**	**40.642**	**5.524**	**299.007**	**0.000**
	오이무침	301	117	89	28	112	72	2.043	1.218	3.429	0.009
	깍두기	303	117	94	23	119	67	2.301	1.334	3.970	0.004
	백미밥	**303**	**117**	**116**	**1**	**148**	**38**	**29.784**	**4.029**	**220.163**	**0.000**

표 3-2-10. 학원에서 발생한 클로스트리디움 감염증의 원인을 찾기 위한 다변량분석 결과

식단 음식물			상대위험비 (오즈비)	95%CI		P
날짜	음식 이름	코드		Lower	Upper	
3/10 저녁	북엇국	A201	0.932	0.382	2.279	0.878
	소불고기볶음	A202	0.911	0.133	6.212	0.924
	미트볼볶음	A203	0.650	0.030	13.943	0.783
	골뱅이무침	A204	1.373	0.420	4.487	0.600
	부추장떡	A205	0.377	0.114	1.248	0.110
	파래무침	A206	1.955	0.852	4.484	0.114
	배추김치	A207	1.247	0.347	4.486	0.735
3/11 저녁	치즈돈까스	B202	0.131	0.008	2.202	0.158
	골뱅이소면	B203	0.843	0.193	3.688	0.821
	무생채	B206	0.644	0.271	1.534	0.321
	깍두기	B207	1.373	0.319	5.914	0.671
	백미밥	B208	12.215	0.336	444.467	0.172
3/12 저녁	오뎅국	C201	1.230	0.417	3.628	0.708
	고등어조림	C203	0.744	0.315	1.756	0.500
	닭볶음탕	**C205**	**35.010**	**2.513**	**487.767**	**0.008**
	오이무침	C206	1.120	0.505	2.483	0.780
	깍두기	C207	0.641	0.177	2.326	0.499
	백미밥	C208	8.315	0.382	181.159	0.178

4 맺음말

역학조사에서 수집한 자료나 관찰역학연구에서 수집된 자료를 이용해서 노출과 질병의 연관성을 분석하고자 하는 경우 교란변수의 효과를 보정하여야 한다. 그렇지 않고 단순한 오즈비나 상대위험도만을 제시할 경우 그것은 왜곡된 수치이기 때문에 잘못된 원인을 찾을 수도 있고, 원인을 못 찾을 수도 있고, 해석이 어려워질 수 있다. 그런데 이러한 개념이 없이 역학조사를 하므로 운이 좋은 경우에는 문제가 없으나 그렇지 않은 경우는 원인을 못 찾는 역학조사가 되고 만다. 여기서는 단순하게 식단만 예로 들었으나, 성별, 연령, 개인위생수준, 사회경제수준 등 다양한 변수들이 연관성에 영향을 미치기 때문에 늘 이를 염두에 두어야 한다. 특수한 상황을 제외하고 오즈비나 상대위험도는 보정된 지표를 제시해야 한다는 것을 명심하자.

또 하나 분명하게 강조하고 싶은 것은 교란변수와 효과수정인자의 구별이다. 이를 구별하지 못하는 경우 효과수정인자를 다변량분석에 넣고 보정하는 상황이 벌어진다. 이렇게 해서 나오는 숫자는 아쉽지만 진실이 아니다. 따라서 다변량분석을 하기 전에 모형에 포함되는 변수들이 효과수정인자인지를 파악하기 위해서 층화분석을 해보는 것이 좋다. 흔히 연령을 교란변수로 많이 취급하지만 감염병의 경우는 연령에 따라 예방접종의 효과가 달라지는 것처럼 연령이 효과수정인자의 역할을 할 때가 많다. 이 부분을 잘 이해하고 보정을 하도록 한다.

현재 역학조사에서 층화분석이나 다변량분석이 잘 안되고 있는 것은 우리나라 역학조사에서 원인을 못찾거나 엉뚱한 것을 원인으로 지목하는 가장 큰 이유 중의 하나이다.

참고 문헌

1. 천병철, 최승아, 유석현 외 역. 중급 역학. 4판. 서울: 범문에듀케이션; 2023.
2. 천병철. 교란인자. 고려대학교 보건학협동과정 역학개론 강의록. 2024.
3. 천병철. 역학조사 방법론. 고려대학교 보건대학원 감염병역학 강의록. 2023.
4. 천병철. 효과수정인자. 고려대학교 보건학협동과정 역학개론 강의록. 2024.
5. 최용석. R과 함께하는 다변량 자료분석. 서울: 경문사; 2018.
6. WHO. Foodborne disease outbreak: Guidelines for investigation and control. 2008.

제3절 유행조사자료를 이용한 감염재생산수 추정법

1 감염재생산수

감염재생산수(reproduction number, R)는 한 명의 감염자가 평균적으로 만드는 이차감염자의 수를 말하며 단위는 명이다. 일부에서 '감염재생산수'나 '감염재생산비'라는 표현을 쓰고 있는데 이는 오류이므로 사용하지 않아야 한다. 감염재생산수는 평균세대기(mean generation time)와 함께 유행곡선을 결정하는 기본 유행모수(epidemic parameter)이므로 유행의 모델링에 있어서 이 두 가지 모수는 필수적으로 이용된다. 이 두 가지 모수에 의해서 유행곡선의 상승 기울기(유행의 성장속도, growth rate)가 결정되므로 사실상 유행곡선을 결정하는 것은 이 두 가지 모수이다. 따라서 어떤 새로운 유행이 있을 때 이 두 가지 모수를 측정하는 것은 매우 중요하며 특히 새로운 감염병이 나타났을 때 이 두 모수를 알아야 앞으로 어떻게 유행이 진행될 것인지를 알 수 있으므로 제일 먼저 측정해야 하는 지표이기도 하다.

새로운 감염병이 나타났을 때는 모든 사람들이 감수성자이므로 이때 측정된 감염재생산수를 기초감염재생산수(basic reproduction number, 약어는 R_0이며 'R naught'라고 읽는다)라고 하고, 유행이 진행되면서 일부 항체를 보유하여 비감수성자가 인구에 생기면 이때는 유효감염재생산수(effective reproduction number, R_e)라고 하는데, 이는 유행이 경과되면서 변하므로 이를 시간의존형 감염재생산수(time-dependant reproduction number, R_t)라고도 한다. 기초감염재생산수는 감염력의 고유한 전파력을 나타내기 때문에 집단면역(herd immunity)을 가질 수 있는 한계치(threshold)를 계산할 때 이용된다. 유효감염재생산수는 백신접종, 사회적 거리두기 등에 따라서 달라지므로 유행 현재 시점에서 향후 유행이 어떻게 진행될지 모델링할 때 유용하다.

예를 들어서 그림 3-3-1과 같이 일차감염자로부터 감염이 전파되는 가상의 상황을 보자. 여기서는 모든 사람이 감수성자라고 가정하자. 일차환자는 A와 B 2명을 감염시켰고, A와 B도 각각 2명씩 감염시켰다. C도 2명을 감염시켰으나 D는 한 명도 감염을 시키지 않았다. 여기서 총감염자는 일차감염자와 A~G까지 총 8명이다. 이 8명이 감염시킨 수는 총 10명이므로 이 유행의 감염재생산수는 식 1과 같이 1.25가 된다. 즉, 1명이 새로운 감염자 1.25명을 만드는 상황이므로 유행은 확산되고 있음을 의미한다.

$$\text{식 1: 기초감염재생자수 } R0=\frac{(2+2+2+2+0+1+0+1)}{8}=\frac{10}{8}=1.25$$

감염재생산수가 1보다 크면 감염자 수가 증가하고 있으며 유행이 확산 중인 것을 의미하며, 1보다 작으면 감염세대가 진행될수록 감염자 수가 증가하여 결국 유행이 종식될 것을 의미한다. 만약 1과 같으면 현재의 감염자 수가 유지될 것이므로 유행이 안정화되는 것을 의미하고 지속적으로 이렇게 유지된다면 풍토병이 된 것을 의미한다.

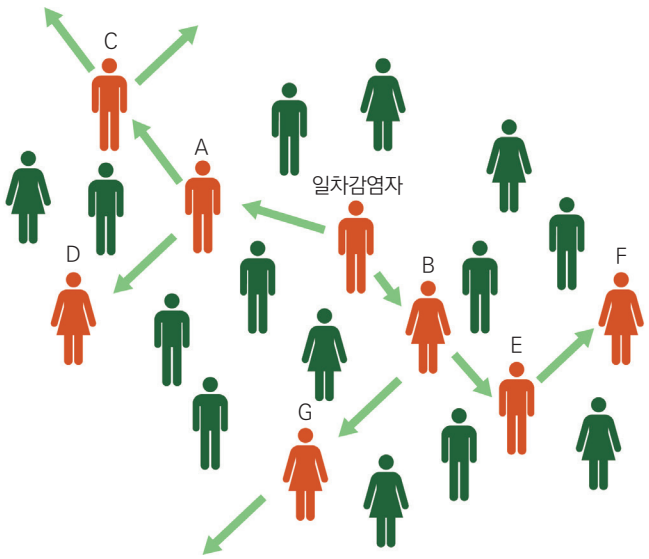


그림 3-3-1. 지역사회 유행에서 감염재생산 수의 계산
전체 8명의 감염자가 총 10명의 감염자를 만들었으므로 감염재생산수는 10/8=1.25가 된다.

2 감염재생산수의 구성요소

감염재생산수는 3가지 요소에 의해서 결정된다. 그림 3-3-2에서 보는 것처럼 감염자가 하루에 몇 명과 접촉을 하는가(κ), 그리고 한번 접촉시 전파확률(β)은 얼마인가, 그리고 감염력을 가진 기간(감염기간, D)은 얼마인가의 3 요소에 의해서 결정된다. 예를 들어서 평균감염기간이 10일이고 하루에 6명과 유효접촉하며 이때 전파확률이 0.05라면 $R=6\times0.05\times10=3$이 된다.

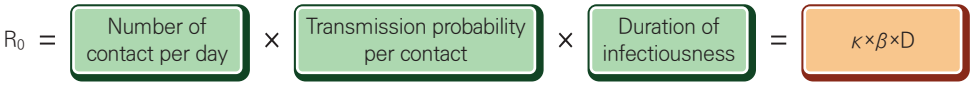


그림 3-3-2. 감염재생산수의 3요소

유행 시 방역조치 중 환자를 격리시키고 학교폐쇄와 같은 사회적 거리두기를 하는 것, 예방접종 등은 바로 κ를 떨어뜨리기 위한 것이며, 마스크 착용이나 손씻기와 같은 전파확률인 β를 감소시키기 위한 것이며 환자를 치료하여 감염기간을 줄이는 것은 D를 감소시키기 위한 것이다. 즉, 우리가 수행하는 모든 방역활동은 위의 카테고리에 들어가며 이는 R을 떨어뜨리기 위한 것이다. 즉, 방역의 목표는 R을 1보다 낮은 수준으로 떨어뜨리는 것이다.

3 감염재생산수의 추정방법

감염재생산수를 추정하는 방법은 매우 다양한데 여기서는 가장 단순한 6개의 방법만 소개한다.

1. 개별 역학조사 자료를 이용하는 방법

만약 누가 누구를 감염시켰는지를 정확하게 역학조사를 통해서 안다면 그림 3-3-1처럼 바로 개별 역학조사자료를 이용해서 R을 계산할 수 있다. 성매개감염병의 경우 접촉여부를 분명히 파악할 수 있으므로 이런 경우에 활용할 수 있으나, 대부분의 호흡기 감염병의 경우는 누가 누구에게 전파되었는지 불분명하고, 잠복기가 길어질수록 이런 감염자-피감염자 접촉의 관계를 파악하는 것은 매우 어렵게 된다. 따라서 현실적으로 접촉여부를 확인할 수 있는 특별한 경우에만 가능하다.

2. 질병의 자연사로부터 추정하는 방법

질병의 자연사(natural history)란 해당 질병의 감염이 일어나는 시점부터 감염이 끝나는 시점까지의 모든 과정을 말한다. 만약 우리가 해당 감염병의 감염기간을 정확히 알고 있고, 한 사람의 감염자의 접촉매트릭스(하루당 만드는 유효접촉의 수)를 측정할 수 있고, 한 번의 접촉 시 전파확률을 알고 있다면 그림 3-3-2의 공식을 이용해서 감염재생산수를 계산할 수 있다. 실제로 SARS가 유행했을 때 홍콩에서

SARS의 R값을 계산할 때 사용했는데, SARS의 D=6일, 접촉시 전파확률 β는 가구내 접촉일 때 0.25, 사회적 접촉일 때 0.1이고 이 둘의 가중치 β=0.149995로 측정되었고, 홍콩인의 평균 접촉인원은 가구내 2명, 사회접촉 12명으로 14명으로 하루당 2.33명이었다(κ=14/6=2.33). 이를 이용해서 감염재생산수를 계산하면 아래와 같이 2.1이 된다.

$$R_0=6\times 2.33\times 0.149995 =2.1$$

이 값은 다른 방법으로 추정한 감염재생산수값과 거의 같았다. 그러나 일반적으로 접촉매트릭스가 정확하게 측정되어 있는 경우가 드물며, 특히 β값은 측정하기가 아주 어려워서 일반적 상황에서 자주 사용되지는 않는다.

3. 유행 발병률로부터 감염재생산수 계산

이 방법은 R값과 발병률(attack rate, A)과의 관계로부터 간단하게 유도되는 식을 이용하는데, 리드-프로스트 모형(Reed-Frost model)으로부터도 유도할 수 있다(식을 유도하는 방법은 천병철, 계량역학강의록, 고려대학교 보건대학원, 2024 참조). 결론적으로 발병률을 A라고 하면 R과의 관계는 다음과 같이 단순화된다.

$$\text{식 2: } R=\frac{\ln(1-A)}{A}$$

예를 들어서 1,000명이 있는 한 지역사회에 인플루엔자가 유행해서 총 400명이 감염되고 유행이 종료되었다고 하자. 이때 발병률 A=400/1,000=0.4가 된다. 이 인플루엔자의 감염재생산수는 식 2에 의해서 다음과 같이 계산할 수 있다.

$$R=\frac{\ln(1-A)}{A}=-\frac{\ln(1-0.4)}{0.4}=-\frac{-0.5108}{0.4}=1.25$$

즉, 평균적으로 1명의 인플루엔자 감염자가 1.28명의 새로운 감염자를 만든 것이다. 이 방법은 간편하고 새로운 감염병에 대해서도 쉽게 구할 수 있다는 장점이 있다. 그렇지만 한 가지 단점은 유행이 모

두 종료된 시점에서야 R값을 구할 수 있다는 것이다. 즉, 유행초기나 유행 중에는 이 방법으로는 R값을 구할 수 없다.

4. 감염자의 평균 연령을 이용한 감염재생산수의 계산

이것은 R이 클수록 접촉이 많거나(κ), 접촉 시 전파확률(β)이 크거나, 감염기간(D)이 길다는 것을 의미하므로 큰 R일수록 감염자의 평균연령이 낮아진다. R이 큰 감염병은 감수성자의 비율을 낮추는데 이를 일찍 달성해 버리기 때문이다. 특히 신종감염병의 경우에는 이를 이용해서 감염자의 평균연령으로 R을 구할 수 있다(식 3, 4).

식 3: $\frac{L}{A}$(기대수명의 구조가 베타분포인 경우)

또는

식 4: $R=1+\frac{L}{A}$(기대수명의 구조가 음이항분포인 경우)

식 3와 식 4에서 A는 감염자의 평균연령이고 L은 해당 지역사회의 평균기대수명이다. 예를 들어서 홍콩의 기대수명이 76세이고 SARS 감염자의 평균연령이 35세라면 SARS의 R값은 76/35=2.1이 된다. 또 다른 예로 메릴랜드의 도시지역의 백일해 환자의 평균연령이 4.9세인 반면, 농촌지역의 감염자의 평균 연령은 6.5세였는데, 이는 같은 백일해라도 도시지역의 R값이 농촌보다 크다는 것을 의미한다. 그러나 이 방법을 이용해서 R값을 추정하기 위해서는 유행이 종료된 이후에 가능하기 때문에 유행 중에는 이 방법으로 R값을 추정할 수 없다.

5. 유행곡선의 성장률을 이용한 방법

이 방법은 유행초기의 유행곡선만으로 감염재생산수를 구할 수 있으므로 실제 상황에서 가장 많이 사용되는 방법이다. 이것은 유행곡선의 성장률(growth rate)이 감염재생산수와 평균세대기에 의해서 결정된다는 것을 이용한 방법이다. 유행곡선의 성장률은 감염자 수가 시간당 얼마나 빠르게 증가하는 것을 나타내는 지표이다. 대부분 지수함수를 따르며 이 방법으로 R을 구하려면 유행곡선으로부터 먼저 성장률을 계산해야 한다. 일반적으로 이 들의 관계는 다음과 같이 정리할 수 있다.

식 5: growth rate(r)≈ln(R)/Tg

식 5에서 Tg는 평균 세대기이며 세대기가 짧을수록 성장률이 급속히 증가하여 감염자 수의 증가가 이루어질 것이라는 것을 알 수 있다. 실제 계절 인플루엔자의 경우 R값은 1.2 정도로 낮지만 세대기가 짧아서 폭발적으로 유행이 일어난다.

위의 식을 R에 대해서 풀면 아래와 같다.

식 6: $R=\exp(\gamma \times T_g)$

예를 들어서 1918년 스페인독감때 미국의 메릴랜드 지역의 유행곡선에서 측정한 성장률은 0.27/일이었고 평균 세대기는 3일이었다. 이것을 식 6에 대입하면 스페인독감의 R값은 다음과 같이 계산된다.

$$R=\exp(\gamma \times T_g)=\exp(0.27 \times 3)=2.2$$

그런데 이 식은 감염자의 평균 감염기간(infectious period)이 3일이라고 할 때, 모든 감염자의 감염기간이 3일까지 지속되다가 3일이 지나면 감염력이 없어지는 가정으로 유도된 식이다. 만약 평균 감염기간은 같은 3일이지만, 시간이 지남에 따라 지수적으로 감소하는 것을 가정하면 R은 다음과 같이 유도된다.

식 7: $R=1+(\gamma \times T_g)$

식 7으로 같은 유행을 계산하면 다음과 같이 R값은 1.81이 된다. 이 값은 이후에 측정된 스페인독감의 R값과 더 가깝다.

$$R=1+(\gamma \times T_g)=1+(0.27 \times 3)=1.81$$

현실적으로 유행곡선에서 성장률을 구하는 것이 쉽지 않아서 (이것은 최대우도함수를 이용해서 구할 수 있지만 복잡하다) 두배증가시간(doubling time, Td)을 구해서 계산하는 경우가 많다. 두배증가시간은 감염자 수가 2배로 늘어나는데 소요되는 평균 시간이다. 이때 r≈ln(2)/Td의 식으로 나타낼 수 있으므로 이 식을 식 4에 대입하면 다음과 같은 관계를 얻을 수 있다.

식 8: $Td \approx \ln(2) \times T_g / \ln(R)$

초기 유행곡선으로부터 R값을 계산하는 방식은 이외에도 여러 가지이지만, 실제로는 R 통계프로그램의 다양한 통계패키지를 이용하면 유행곡선으로부터 다양한 방법으로 감염재생산수를 쉽게 구할 수 있다.

여기에서 소개한 방법 이외에도 감염재생산수를 추정할 수 있는 방법은 매우 다양하다.

4 R_0 통계패키지를 이용한 감염재생산수 추정법

R 통계프로그램은 감염재생산수나 세대기를 추정할 수 있는 다양한 패키지(package)를 제공하고 있는데 이를 이용하면 가장 쉽게 유행곡선으로부터 이러한 유행모수를 추정할 수 있다. R 통계프로그램에 이러한 유행모수를 측정하는데 사용하는 패키지로는 R0, EpiEstim, EpiNow2, surveillance 등이 대표적이다(surveillance 패키지는 제2권 제2장 참조). 여기서는 R_0 패키지를 이용한 방법을 간단히 설명한다. 이를 이용하기 위해서는 먼저 R 통계프로그램을 구동한 후 R_0 패키지를 선택하여 설치한다.

1. R_0 패키지를 이용하여 발병률로부터 감염재생산수 구하기

R_0 패키지는 발병률을 알고 있을 때 이로부터 감염재생산수를 구하는 함수를 제공한다. 예를 들어서 영국의 한 지방에서 1957년 아시아독감이 유행 시 전체 9,800명 중에서 3,303명(발병률: 3,303/9,800=0.337)이 독감에 걸렸다. 이를 이용해서 아시아독감의 R_0를 구하는 코드는 박스 3-3-1과 같다. 표 3-3-1에서 붉은색은 입력한 R 명령어 프로그램이고 검은색은 실행시킨 후 나오는 계산값이다.

박스 3-3-1. 발병률로부터 감염재생산수를 구하는 R 코드(R_0 패키지 이용)

```
> library (R0)
> est.R0.AR(pop.size=9800, S0=1, incid=3303)                    # 명령어 1
Reproduction number estimate using Attack Rate method.
R: 1.219561[1.211613 , 1.227655]

> est.R0.AR(pop.size=9800, S0=1, AR=0.337)                      # 명령어 2
Reproduction number estimate using Attack Rate method.
R : 1.219526[1.211579 , 1.227619]
```

명령어 1은 발병률(attack rate, AR) 방법으로 R_0를 구하라는 함수이며, 최초 감염자는 1명(S_0=1), 유행종료후 감염자는 3,303명이라는 의미이며, 명령어 2는 바로 발병률(AR)을 입력한 것이다. 이 두 상황은 같기 때문에 같은 R과 R의 95% 신뢰구간이 산출되었다. 소수점자리에서 약간 차이가 나는 것은 발병률 입력 시 소수점을 3자리에서 끊었기 때문이다.

2. R_0 패키지를 이용하여 유행곡선으로부터 감염재생산수 구하기

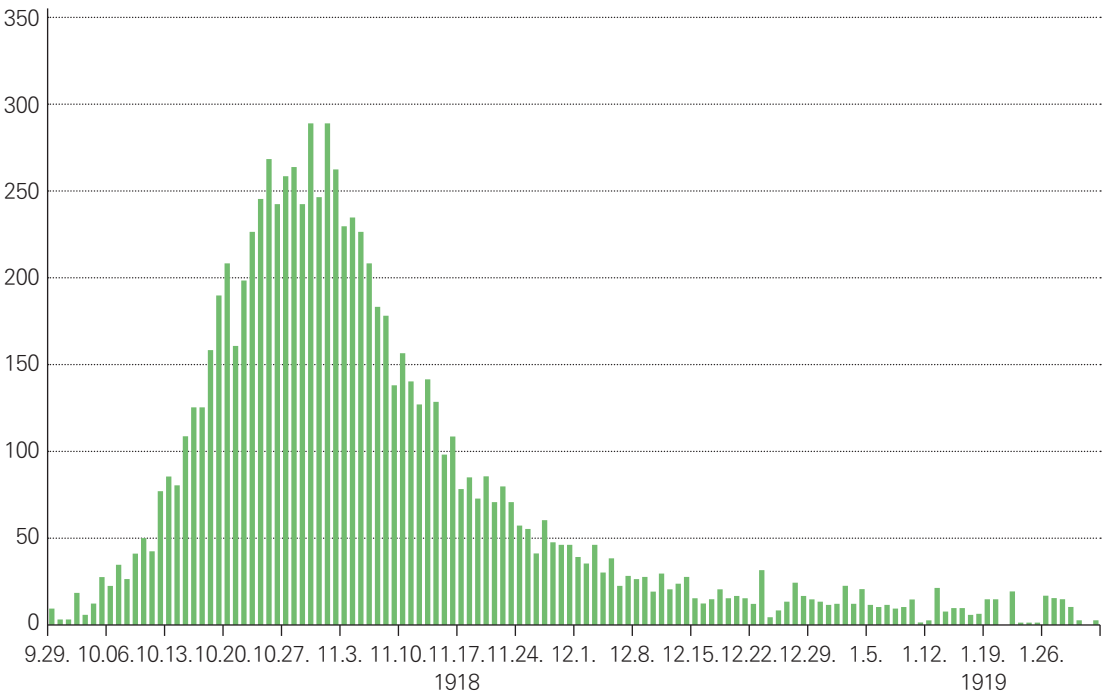


그림 3-3-3. 1918년 스페인독감 시 독일 한 지역의 독감유행곡선

그림 3-3-3은 R_0패키지에 내장되어 있는 자료를 이용해서 그린 유행곡선으로 독일의 한 지역에서 1918년 스페인독감 발생자료이다. 이 유행곡선을 보면 1918년 9월 말부터 환자가 발생하기 시작해서 11월 초에 정점을 찍고 이후 사례수가 감소하였다. 이 유행자료를 분석하면 첫 환자 발생 이후 27일간 사례 수가 지수적으로 증가하였는데 이 구간의 증가율이 유행곡선의 성장률(growth rate)이 된다. R_0 패키지를 이용하면 우리가 성장률을 구할 필요 없이 패키지에서 계산해 주기 때문에 이를 이용한 감염재생산수를 구하는 명령어만 입력하면 된다. 이 유행의 전체 유행기간은 126일이다. 이 자료를 이용해서 감염재생산수를 구해 보자(박스 3-3-2).

R_0패키지에서는 유행곡선에서 감염재생산수를 구하는데 3가지 방법을 제시하고 있는데 하나는 유행곡선 성장률법(EG), 최대우도법(ML), 연속베이지안 추정법(SB)이다. 그리고 유행시간에 따라 변하는 Rt 값을 추정하는 명령어도 있다. 이들의 R 명령어와 결과는 박스 3-3-2와 같다.

박스 3-3-2. 유행곡선으로부터 기초감염재생산수(R_0)를 구하는 R 코드(R_0 패키지 이용)

```
> library (R0)
> data (Germany.1918)                                  # 사용할 자료명 지정
> mGT ← generation.time (“gamma”, c(3,1.5))            # 평균세대기 지정
> est.R0.EG (Germany.1918, mGT, begin=1, end=27)       # 명령어 3
Reproduction number estimate using Exponential Growth method.
R : 1.525895[ 1.494984 , 1.557779 ]

> est.R0.ML (Germany.1918, mGT, begin=1, end=27)       # 명령어 4
Reproduction number estimate using Maximum Likelihood method.
R : 2.011089[ 1.902864 , 2.123272 ]

> est.R0.SB (Germany.1918, mGT, begin=1, end=126)      # 명령어 5
Reproduction number estimate using Sequential Bayesian method.
0 0 3.43 0.31 1.4 2.67 1.86 2.26 1.68 2.04 ···
```

명령어 3의 est.R0.EG (Germany.1918, mGT, begin=1, end=27)는 지수적 성장률(EG)을 이용해서 R_0를 추정하되 Germany.1918 자료를 이용하고 평균 세대기는 mGT를 이용하며 유행의 첫째 날에서 27일째의 자료를 이용하라는 것이다. 여기서 평균 세대기는 윗줄에 정의하여 두었는데 대부분 감마분포에 적합한 자료를 사용하기 때문에 gamma분포로 지정되어 있다. 많은 경우 세대기가 알려져 있어서 이를 활용하지만, 만약 해당 질병의 세대기가 알려져 있지 않다면 유행곡선에서 세대기를 추정하면 된다. 이 방법도 여러 가지가 있으나 R에서는 fitdist 함수나 EpiEstim 패키지를 쓰면 편리하다. 첫째 날에서 27일까지의 자료를 쓰는 이유는 이 구간이 성장구간이기 때문이다. 이 자료를 로그 그래

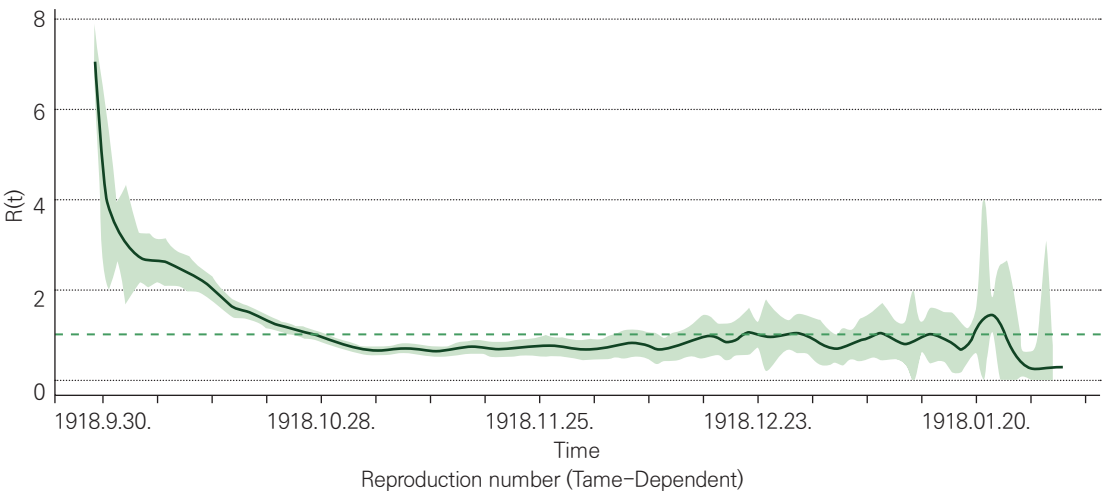
프로 다시 그리면 이 부분은 직선이 되고 이 기울기가 성장률에 해당한다. 이 지점은 유행곡선에 찾을 수 있다. 명령어 3에 의해서 R_0는 1.53과 95%신뢰구간이 제시되어 있다.

명령어 4는 최대우도법(ML)을 이용하는 방법으로 명령어 구조는 앞의 명령어 3과 같고, EG 대신 ML이 들어가 있다. 이 방법으로 구했을 때 R_0가 2.0으로 커진 것을 알 수 있다. 명령어 5는 연속베이지안 추정법이고 매일매일의 R값 추정치를 보여준다.

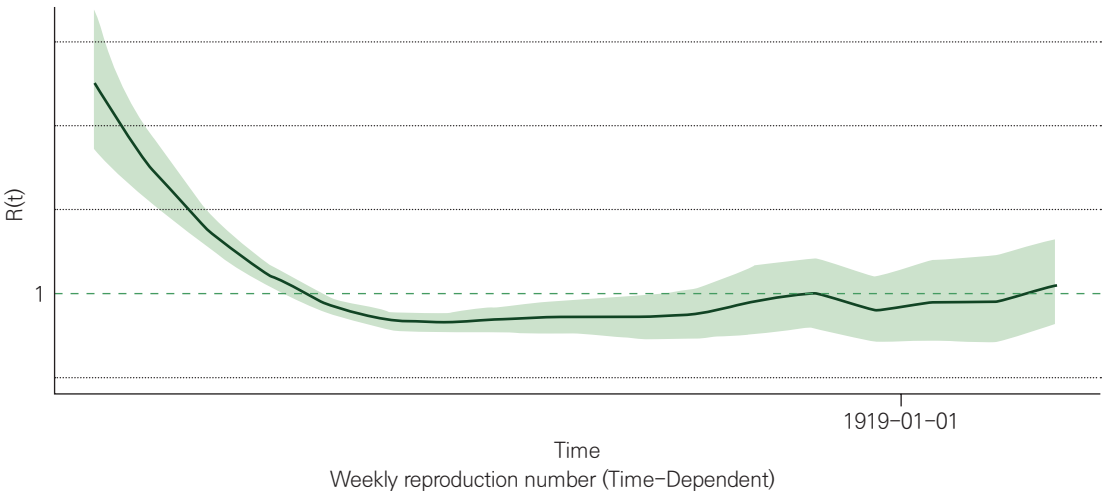
여기까지가 R_0를 추정하는 방법이라면 박스 3-3-3은 유행기간 중 유효감염재생산수(Rt)가 시간에 따라 변화한 것을 보여준다. R_0패키지에서 이에 해당하는 함수는 TD이다. 박스 3-3-3의 명령어 6은 1일에서 126일간 매일 유효감염재생산수를 구하라는 함수이며, 이에 처음에 7.0부터 시작해서 계속 떨어져서 27일 정도에 1로 수렴하고 이후에 1을 사이에 두고 오르락내리락 하는 Rt값을 확인할 수 있다. 이는 명령어 7에 의한 그래프에서 명료하게 볼 수 있다. 이렇게 매일매일의 값이 너무 변동이 크므로 이를 주간 단위로 바꾼 것이 명령어 8이고 그 결과 주별 Rt 값은 3.5에서 시작해서 훨씬 안정적으로 추산되었다. 이 주간 Rt값과 95%신뢰구간도 그래프로 볼 수 있다(박스 3-3-3).

박스 3-3-3. 유행곡선으로부터 유효감염재생산수(Rt)를 구하는 R 코드(R_0 패키지 이용)

```
> TD <- est.R0.TD (Germany.1918, mGT, begin=1, end=126, nsim=1000)
> est.R0.TD (Germany.1918, mGT, begin=1, end=126, nsim=1000)        # 명령어 6
Reproduction number estimate using Time-Dependent method.
7.038388 4.350397 3.777973 3.310597 2.997636 2.803256 2.67985 2.62434 2.631525
> plot (TD)                                                         # 명령어 7
```



```
> TD.weekly <- smooth.Rt (TD,7)                                    # 명령어 8
> TD.weekly
Reproduction number estimate using Time-Dependent method. 3.515032 2.450455 1.698659 1.191296 0.8377125
0.6685603 0.6552762 0.6972486 0.7204385 0.7163884 ...
> plot (TD.weekly)
```



5 맺음말

여기서는 R_0와 Rt를 포함한 R값을 추정하는 원리와 방법을 간단히 살펴보았다. R값과 세대기는 어떤 유행이든지 기본적으로 추정해야 할 기본 유행모수이다. 비단 이들을 추정하는 방법과 함께 이 두 가지 모수의 의미를 명확하게 이해해야 할 필요가 있다. 감염재생산수와 세대기는 지역사회 감염병 유행을 이해할 때 필수적인 개념이며 특히 새로운 감염병이 나타났을 때 이들 유행모수를 먼저 구할 수 있어야 한다.

R 통계프로그램의 R_0패키지나 EpiEstim 패키지를 이용하면 이들을 쉽게 구할 수 있으며 더욱이 이들 패키지 사용법은 과거처럼 다 외우지 않아도 인공지능프로그램의 도움을 받아서 R 명령어를 쉽게 익힐 수 있는 시대가 되었고, 유튜브나 인터넷의 다양한 강좌가 개설되어 있기 때문에 참고문헌과 함께 이를 활용해서 구체적 방법을 더 익혀서 실제 자료에 활용했으면 한다.

참고 문헌

1. 천병철. 감염병 유행지표의 산출. 한국역학회 하계연수자료집. 2021.
2. 천병철. 유행모수의 추정. 고려대학교 보건대학원 계량역학 강의록. 2024.
3. Bettencourt LM, Ribeiro RM. Real time bayesian estimation of the epidemic potential of emerging infectious diseases. PLoS One. 2008;3(5):e2185.
4. Dietz K. The estimation of the basic reproduction number for infectious diseases. Stat Methods Med Res. 1993;2(1):23-41.
5. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc Biol Sci. 2007;274(1609):599-604.
6. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol. 2004;160(6):509-16.
7. White LF, Wallinga J, Finelli L, et al. Estimation of the reproductive number and the serial interval in early phase of the 2009 influenza A/H1N1 pandemic in the USA. Influenza Other Respir Viruses. 2009;3(6):267-76.

R_0패키지 이외에 EpiEstim과 EpiNow2 패키지 자료

1. EpiEstim: https://cran.r-project.org/web/packages/EpiEstim/vignettes/demo.html.
2. EpiNow2: https://doi.org/10.12688/wellcomeopenres.16006.2.

제4절 수인성 및 식품매개감염병 유행 시 역추적 조사와 환경 평가의 활용

수인성 및 식품매개감염병 유행 역학조사는 조사 목적으로 원인병원체 규명, 감염원 규명을 목표로 한다. 감염원을 판단할 때 질병관리청의 수인성 및 식품매개감염병 관리 지침에서는 역학적 근거(통계적 유의성 포함)와 실험실 근거(보존식 검사 결과)를 기준으로 주로 판단하고 있다. 미국 질병통제예방센터의 경우 감염원을 판단할 때 역학적 근거, 실험실 근거, 역추적 조사의 근거라는 3가지를 기준으로 판단하여 국내와는 약간의 차이가 있다.

역추적 조사(traceback investigation)는 감염원의 생산과 유통 과정을 확인하여 어느 단계에서 오염되었는지를 확인하고자 하는 조사이다. 보통 여러 유행이 발생하였는데 원인병원체가 동일하여 하나의 유행으로 묶을 수 있는 경우에 시행할 수 있다. 이때는 각 음식을 만드는 장소에서 오염된 것이 아닌 음식의 식재료가 이미 오염된 상태로 도착하였을 가능성이 높다. 이를 조사하기 위해서는 여러 유행에서 동일한 식재료를 공급받았는지를 확인할 필요가 있으며, 각 식재료별 소매상과 도매상의 공통점을 확인할 수 있다. 공통되는 장소를 찾았다면 그 이전 농장도 동일한지를 확인할 수 있으며, 공통되는 한 장소를 오염된 지점으로 추정하여 역추적 조사를 마무리할 수 있다.

환경 평가(environmental assessment)는 의심하는 감염원을 확인한 후 시행할 수 있으며, 의심하는 음식이 어떻게 오염되었는지와 왜 오염되었는지를 규명하고자 시행한다. 특정 음식이 어떻게 오염되었는지는 기여 요인(contributing factor)을 찾는 과정으로 음식을 만드는 과정과 보관하는 과정을 파악하여 원인병원체의 오염, 생존, 증식을 조사한다. 특정 음식이 왜 오염되었는지는 유행을 막고 예방하기 위한 근본 원인(root causes)을 찾는 과정으로 사람(위생 문화 결여), 장비(부족한 용량), 경제(대체인력 부재), 과정(표준과정을 따르지 않음), 음식(유통기한 초과)과 같은 5가지로 구분하여 파악할 수 있다. 여기에서는 환경 평가 중 어떻게 오염되었는지를 중심으로 기술한다.

국내에서는 여러 곳에서 발생한 유행일 때 역추적 조사를 시행할 때가 있으며, 이는 질병관리청의 협조로 식품의약품안전처에서 주로 시행한다. 환경 평가는 모든 음식을 대상으로 할 수 없기에 의심되는 감염원이 있을 경우에 시행할 수 있지만 국내에서는 환경 평가를 시행하는 경우가 드물다. 하지만 감염원이 밝혀지면 기여 요인(오염, 병원체 생존, 증식)을 살피기 위해 보건소 내 위생 관련 부서에 요청하여 감염원을 만드는 과정을 재연할 수 있으며, 필요하다면 역학조사자가 직접 요청할 수도 있다.

1 역추적 조사

역추적 조사는 관련된 식재료를 광범위하게 찾아 공통점을 통해 오염된 장소를 찾아낼 수 있으며, 크게 3가지 단계로 이야기할 수 있다. 첫 번째 단계에서는 사례에게 언제 어디에서 음식을 구매하였는지를 확인하고, 음식을 만든 곳을 확인한다. 이 과정에서 구입한 식재료의 제품번호(lot number)나 제조업체의 이름을 알면 좀 더 수월할 수 있다. 두 번째 단계에서는 구입한 소매점이나 음식점을 확인하고, 이후 도매점과 같은 배급업체를 확인한다. 배급업체는 도매점과 수입업체와 같은 식재료 유통 과정을 포함한 전 과정을 의미하며, 여러 단계의 배급업체가 존재할 수 있다. 세 번째 단계에서는 마지막인 생산업체를 확인하며, 생산업체가 만든 물품 중 언제, 어디에서 만든 것(혹은 제품번호)이 문제인지를 찾아낼 수 있다. 이러한 역추적 조사로 문제가 되는 음식의 식재료가 오염된 곳을 찾아낼 수 있다. 이후에는 추가로 오염이 된 장소에서 어떠한 과정을 통해 오염이 되었는지(예로 바닷물 세척으로 인한 파래 오염 가능성)를 밝혀내고자 할 수 있으며, 이는 환경 평가에 해당하는 부분이다.

경기도 고양시의 A음식점에서 주로 김밥을 섭취한 사람들에서 살모넬라증이 발생하였다고 신고되었다. 한 명의 사망 사례가 언론에 보도되었고(최종적으로는 살모넬라증과 관련이 없는 것으로 봄), 이후 사례 찾기를 통해 106명의 사례를 발견하였다. 사례들이 섭취한 음식과 물을 모두 조사한 결과 김밥을 섭취한 사람이 84.9%로 가장 높았으며, 다양한 김밥(참치, 일반, 웰빙, 김치김밥 등)을 섭취하였다. 며칠 후 고양시의 B음식점에서 웰빙김밥을 주로 섭취한 5명의 살모넬라증 사례가 신고되었다. 한 시에서 수인성 및 식품매개감염병 유행이 며칠 사이에 2건이 신고된 것은 드문 일이며, 또한 같은 살모넬라증으로 신고되는 일은 더욱 드물다. 이 2건의 유행에서 공통점이 있을 수 있어 공통점을 찾기 위한 역추적 조사를 시행하였다.

A음식점에서 사례들이 다양한 김밥을 섭취하였는데, 공통되는 식재료는 계란(지단), 우엉, 단무지, 당근, 상추, 어묵이었다. B음식점에서 사례들이 섭취한 김밥은 웰빙김밥이며, 계란(지단), 게살, 양상추, 아삭이고추, 당근, 우엉을 식재료로 사용하였다. 식품의약품안전처에서 A음식점과 B음식점에서 사용한 식재료의 역추적 조사를 실시하였고 결과는 표 3-4-1에 제시하였다. A음식점과 B음식점에서 공통된 식재료는 계란, 우엉, 당근이며, 도매점이 겹치는 곳은 계란밖에 없었고 이 계란은 Z농장에서 공급받았다. 역추적 조사 결과에서 Z농장을 공통점으로 찾았으며 계란이 오염된 곳은 닭을 키우고 있는 Z농장으로 추정하였다.

표 3-4-1. 두 음식점의 식재료 역추적 조사 결과

소매점	도매점	원산지	식재료	취급 기간
A 음식점				
V 상회	W상회	Z농장	계란	8월 20~24일
D할인매장	H유통	미국	양상추	8월 21~23일
E야채청과	J유통	국내	양상추	모름
	J농산	국내	아삭이고추, 청양고추 등	모름
F 식품	K유통	국내	깻잎, 로메인 청상추	모름
	L유통	중국	당근	모름
	M종합상사	국내(서울 우유)	체다 치즈	8월 23일
B 음식점				
V 상회	W상회	Z농장	계란	8월 20~24일
G 식자재마트	N농산	국내	양상추, 피망, 고추, 파프리카	8월 16~23일
	O유통	국내	상추, 깻잎, 당근	8월 16~23일
	P농산	국내	상추, 깻잎	8월 16~23일
	O유통	국내	깻잎	8월 16~23일
	R농산	국내 및 해외	당근	8월 16~23일
	S농산	국내	당근	8월 16~23일
	T푸드	중국	우엉	8월 16~23일

역추적 조사를 시행한 뒤에 A음식점과 B음식점의 사례들에서 검출된 살모넬라균을 비교한 결과 실험실적으로 동일하다는 것을 확인하여 이 두 유행은 하나로 묶인다는 것을 다시 확인하였다. W상회를 통해 계란을 공급받은 곳은 13개 음식점이었으며, A음식점과 B음식점을 제외한 11개 음식점에서 추가 수인성 및 식품매개감염병 유행 발생 보고는 없었다. A음식점은 8월 20일에 들어온 계란과 23일에 들어온 계란을 사용하여 지단을 만들었다. B음식점은 8월 20일에 들어온 계란을 사용하였고, 두 음식점 모두 공통적으로 8월 20일에 들어온 계란을 사용하였다. 역추적 조사를 통해 Z농장을 오염된 지점으로 추정하였고, 특히 8월 20일에 공급된 계란을 의심하였다. 따라서 추가로 Z농장을 방문해서 주로 8월 20일 이전의 계란과 닭에 대해서 물어보고(인터뷰), Z농장에서 환경 검체(닭 분변, 계란, 주변 계사, 바닥의 흙 등)를 채취할 수 있었다. 이 환경 검체에서 사례에서 발견된 살모넬라균과 동일한 병원체가 검출된다면 Z농장을 통해 유행이 발생한 것을 확실하게 밝혀낼 수 있었고, Z농장의 닭을 관리하여 추가 유행 발생을 막으려는 목적을 달성할 수도 있었다. 하지만 Z농장은 관할 부처가 농림축산식품부로 질병관리청이나 식품의약품안전처에는 권한이 없었다. 관련 기관에 조사를 요청하였지만 Z농장의 추가 조사를 시행할 수 없는 아쉬움이 있었다. 이처럼 식품과 관련된 역추적 조사는 식품의약품안전처,

농림축산식품부 등 관련 행정부서의 협조가 필수적이다.

2 환경 평가

한 장소에서 섭취한 공통점이 있는 유행은 해당 장소에서 오염이 발생하였다는 것을 시사한다. 해당 장소의 특정 음식을 감염원으로 의심할 경우 조사자는 조리종사자(또는 음식을 만든 사람)와 인터뷰를 하여 어떤 식재료를 사용하였는지, 음식을 만드는 과정은 어떤지, 음식을 만들거나 보관할 때 온도는 얼마인지와 같은 정보를 파악한다. 조리 장소의 청결도나 조리종사자의 건강 상태를 확인하며, 이러한 환경 평가를 통해 유행을 막고 추가 발생을 예방할 수 있는 방법을 찾을 수 있다. 여기에서는 유행 역학조사의 기여 요인을 찾는 과정을 중심으로 기술하며, 근본 원인은 식품관련 전문가가 다룰 수 있는 부분으로 보아 더 이상 기술하지 않았다.

기여 요인은 오염(contamination), 병원체의 생존(survival), 증식(proliferation)의 3가지로 구분하여 파악할 수 있다. 오염은 원인병원체가 어떻게 해당 음식으로 들어갈 수 있는지를 파악하는 부분이다. 예컨대, 아픈 조리종사자를 통해 음식으로 들어간 것으로 본다면 음식을 다룰 때 맨손으로 하였는지, 일회용 장갑을 썼는지와 같은 부분을 주로 볼 수 있다. 식재료를 다룰 때 교차 오염이 된 것으로 본다면 식재료마다 각기 다른 도마를 썼는지, 도마를 쓰고 나서 잘 씻었는지, 식재료를 다루고 나서 손을 잘 씻었는지와 같은 부분을 확인할 수 있다. 생존은 원인병원체인 미생물이 음식을 요리할 때 살아남을 수 있는지를 파악하는 부분이다. 예컨대, 음식을 오래 익혔는지(조리 시간 파악)와 충분히 익혔는지(온도 확인)를 들 수 있다. 증식은 조리하고 난 후 음식에 있는 원인병원체가 증식할 수 있는가를 파악하는 부분이다. 이 경우 세균은 증식할 수 있지만 바이러스는 음식에서 증식할 수 없다는 사실을 염두에 두고 파악해야 한다. 온도에 따라 각 세균이 증식할 수 있는 능력이 다르며 보통 실온에 보관한 경우에는 냉장고에 보관할 때보다 위험한 것으로 볼 수 있다. 또한 병원체마다 감염량(감염에 필요한 용량, infectious dose 50%)이 다르며, 감염량이 큰 경우에는 증식의 과정이 더욱 필요하다고 볼 수 있다.

환경 평가에서 특정 음식이 어떻게 오염되었는지(기여 요인)를 찾고자 할 때는 조리종사자와의 인터뷰를 통해 전반적인 내용을 파악할 수 있다. 또한 가능하다면 특정 음식을 만드는 과정을 직접 재연하여 시간의 흐름에 따라 정리하여 오염, 생존, 증식이라는 점을 생각해 볼 수 있다. 앞의 두 음식점의 살모넬라증 유행에서는 초반에 감염원으로 김밥과 김밥에 사용한 계란으로 만든 지단을 의심하였다. A음식점(106명 발생)은 8월 20일에 받은 계란을 외부의 상온에 보관하였고, 지단을 8월 22일 오후 5시에

대량으로 만들고 다음날에 사용하였다. 지단을 만들고 김밥을 만드는 과정을 재연할 때 맨손으로 계란을 깨고 휘저은 후 계란을 부쳤다. 계란을 부칠 때 얇은 팬을 사용하여 앞면을 구운 후 뒤집고 나서 수초 만에 꺼내어 뒷면은 충분히 익지 않았다. 부친 계란을 잘라서 지단을 만들고 김밥 조리대 아래 냉장고에 지단을 보관하였다. 야채를 써는 도마를 확인한 결과 노랗게 변한 부분이 있어 이 도마를 사용해서 지단도 썰었을 가능성이 있었다. 김밥을 만들 때 지단을 스테인리스 보관함에 일정량을 넣어두고 사용하며, 떨어지면 다시 보충하였다. 김밥을 만들 때는 위생장갑을 끼고 각종 식재료를 올려서 만들며, 위생장갑은 하루에 3번 정도만 교체하였다.

B음식점(5명 발생)은 8월 20일 12시경에 계란을 받았고 바로 냉장고에 보관하였다. 당일 새벽 5~6시경에 지단을 만들고 나서 냉장고에 보관하였다. 재연 과정에서 라텍스 장갑을 끼고 그 위에 일회용 장갑을 끼고 계란을 깼다. 계란을 부칠 때는 라텍스 장갑을 끼고 있었고, 좀 더 깊은 팬을 사용하였고 뒷면은 어느 정도 익혔다(가운데 온도는 확인하지 못함). 부친 계란을 잘라서 지단을 만들고 업소용 냉장고에 지단을 보관하였다. 김밥을 만들 때 지단을 스테인리스 보관함에 일정량을 넣어두고 사용하였으며, 떨어지면 다시 보충하였다. 김밥을 만들 때는 위생장갑을 끼고 각종 식재료를 올려서 만들었으며 위생장갑은 수시로 교체하였다.

이 두 음식점은 원재료인 계란이 이미 오염되었기에, 생존과 증식과 관련한 요인을 주로 찾아볼 수 있다. 생존의 경우 지단을 만들 때의 과정을 주로 보게 되며, A음식점은 계란을 부칠 때 뒷면을 충분히 익히지 않아 생존할 수 있는 가능성이 높았으며, B음식점은 두껍게 부쳐서 속까지 다 안 익었을 가능성이 있었다. 증식의 경우 만들고 난 후의 보관 과정을 주로 보게 되며, A음식점은 냉장고에 보관하였고 B음식점도 냉장고에 보관하였다. 다만 A음식점은 전날 미리 만들어 두었고 B음식점은 당일 새벽에 만들어 보관 시간의 차이가 있고 증식 정도에도 차이가 있었을 가능성이 있었다. 또한 증식과 관련한 요인으로 계란의 외부 보관을 고려할 수 있는데, A음식점은 생계란을 상온에서 보관하였지만 B음식점은 냉장고에 보관하여 요리하기 이전 증식 정도에도 차이가 있었을 수 있다. 교차 오염의 경우 A음식점은 계란을 깰 때 맨손으로 하였고 지단을 만들 때 쓰는 도마를 야채를 썰 때도 사용할 가능성이 있었다(A음식점 사례 모두가 김밥을 먹은 것은 아님). 이를 정리하여 표 3-4-2에 제시하였고, 추가로 필요한 부분이지만 조사하지 못한 항목(가운데 온도, 상온에 두는 시간)은 모름으로 기술하였다.

표 3-4-2. 두 음식점의 환경 평가 및 재연 결과

구분	A음식점(106명 발생)	B음식점(5명 발생)
오염		
계란 자체의 오염	동일 농장의 오염된 계란	동일 농장의 오염된 계란

구분	A음식점(106명 발생)	B음식점(5명 발생)
계란의 교차 오염	맨손으로 계란을 깸	일회용 위생장갑을 끼고 깸
도마의 교차 오염	도마를 같이 사용했을 가능성	찾기 어려움
생존		
지단의 두께	얇음	두꺼움
지단을 익히는 시간	짧음(뒷면을 익히지 않음)	길(뒷면도 어느 정도 익힘)
지단의 가운데 온도	모름	모름
증식		
지단을 만든 후 보관 온도	냉장고에 보관	냉장고에 보관
지단을 만든 후 보관 기간	길(전날 만듦)	짧음(당일 새벽에 만듦)
지단을 상온에 두는 시간	모름	모름
계란의 보관 온도	상온에 보관	냉장고에 보관

환경 평가는 음식점이 아닌 식재료를 만든 회사에서도 시행할 수가 있으며, 파래로 인한 노로바이러스 유행 사례를 들 수 있다. 이 두 초등학교에서 파래로 만든 음식으로 인한 노로바이러스 유행이 최초로 발생하였다. 두 파래로 인한 음식은 파래무생채와 파래배무침으로 둘 다 익히지 않은 음식이었다. 파래무생채와 파래배무침을 공급받은 곳은 동일한 X회사였으며, 이 회사를 방문하여 파래를 납품할 때까지의 과정을 시간의 흐름에 따라 확인하였다. 파래 원료는 완도와 부산 인근 파래 양식장에서 구입하였으며, 실온 상태로 일반 작업장 바닥에 불결한 상태로 보관하였다. 부산 인근 바닷물을 이용하여 1차로 세척 후 소분하여 트레이에 보관하여 각 초등학교에 납품하였다. 원인병원체가 노로바이러스로 증식의 개념이 아니라 오염을 중심으로 살펴볼 수 있으며, 크게 파래 원료가 오염되었을 가능성과 바닷물로 세척하면서 오염되었을 가능성으로 나눌 수 있었다. 환경 검체를 채취한 조사를 시행하기로 하였으며, 세척 전 파래 원료(완도와 부산 인근 파래 양식장), 세척 후 파래, 세척수인 바닷물, 인근 바닷물에서 검체를 채취하였다. 노로바이러스 검사를 시행한 결과 모두 노로바이러스가 검출되었다. 이 중에서 세척 후 파래와 세척수인 바닷물의 노로바이러스가 유전자 계통분석에서 하나로 묶였으며(98% 상동성), 바닷물로 세척하는 과정에서 파래가 오염된 것으로 추정하였다. 이와 같이 환경 평가 중에도 과정을 확인한 후 각 단계마다 적절한 환경 검체를 채취하면 오염 과정을 밝힐 수 있는 근거를 찾을 수 있다.

참고 문헌

1. 질병관리청. 2024년도 수인성 및 식품매개감염병 관리지침. 2024. 68쪽.
2. Centers for Disease Control and Prevention [Internet]. Environmental assessments [Cited 2024 Jan 15]. Available

from: https://www.cdc.gov/restaurant-food-safety/php/investigations/environ-assess.html.

3. Eun JS, Han J, Lim JH, et al. Salmonellosis outbreaks linked to eggs at 2 gimbap restaurants in Korea. Epidemiol Health. 2024;46:e2024036.
4. MacDonald PDM. Methods in field epidemiology. MA: Jones & Bartlett Learning; 2011. pp. 243-9.
5. Park JH, Jeong HS, Lee JS, et al. First norovirus outbreaks associated with consumption of green seaweed (*Enteromorpha* spp.) in South Korea. Epidemiol Infect. 2015;143(3):515-21.
6. Ramussen SA, Goodman RA. The CDC field epidemiology manual. New York: Oxford University Press; 2019. pp.428-33.

역학조사관 교육
표준교재

Textbook For
Korea Field Epidemiology
Training Program





4장

감염병진단 역량강화







제1절 감염병 진단의 정확성과 신뢰도에 대한 이해
제2절 차세대염기서열 기반 유전체분석검사의 이해

집필자 이관, 조민철
검토자 박재선, 이현정

제1절 감염병 진단의 정확성과 신뢰도에 대한 이해

1 개요

코로나19 대유행을 겪으면서 역학조사와 방역 부문에서 적용된 다양한 첨단 기술은 유행 관리와 종식에 큰 역할을 하였다는 것은 분명하다. 역학조사를 하면서 동선을 파악하는 데 사용된 역학조사지원시스템(Epidemic Investigation Support System, EISS), 감염 예방과 사망자 감소에 이바지한 백신과 치료제 등이 있었다. 이 중에서 무엇보다도 코로나19 유행 관리에 큰 역할을 하였던 것으로 코로나19를 신속하게 진단할 수 있는 진단 시스템을 들 수 있다. 실시간 역전사중합효소연쇄반응 검사(Reverse Transcription Polymerase Chain Reaction, RT-PCR)에서부터 신속항원검사(Rapid Antigen Test, RAT)까지 다양한 진단검사 도구가 코로나19를 진단하는 데 이용된 바 있다.

특히 유행 초기 백신도 치료제도 없던 시기에 도입된 RT-PCR 검사는 대규모 확산이 이루어지기 전에 확진자를 신속하게 발견하고, 격리, 치료할 수 있었던 밑바탕이 되었다. 하지만 이러한 진단검사 도구가 개발되거나 현장에서 적용할 때 민감도, 특이도, 양성예측도, 유병률 등 검사의 타당도(정확성)와 신뢰도(정밀성)에 대한 다양한 역학 용어를 잘 이해해야 한다. 비단 코로나19 유행 시뿐만 아니라 평소에도 타당도와 신뢰도는 많이 사용되는 개념이다. 법정감염병 감시체계를 평가할 때 얼마만큼 감시체계를 통해 법정감염병을 잘 찾아내는가를 볼 수 있는 지표가 바로 감시체계의 민감도이다.

결국, 국내 법정감염병 감시체계는 의료기관의 신고를 통해 운영되는 만큼 기준점을 건강보험청구자료로 두고, 감시체계로 얼마만큼 신고가 되었는지를 보게 되면 이게 바로 감시체계의 정확성을 판단할 수 있는 민감도 지표이다.

역학조사관은 질병 감시, 유행 조사, 공중보건 연구 등을 수행하므로, 데이터를 해석하고 결론을 내릴 때 측정 도구의 정확성(타당도)과 신뢰도(정밀성)를 잘 이해해야 한다. 이 절에서는 검사의 타당도와 신뢰도의 개념 및 이들이 실제 현장에서 어떻게 활용되고 있는지 살펴보기로 하자.

2 타당도와 신뢰도의 개념

타당도(validity)는 정확도(accuracy)라고 하기도 하고, 측정하고자 의도한 참값과의 일치하는 정도를 이야기하고, 신뢰도(reliability)는 정밀도(precision)라고 하기도 하며, 참값과는 무관하게 측정 간에 일치하는 정도를 이야기한다. 활을 쏠 때 과녁의 중심에서는 벗어나 있지만 일관성 있게 모여 있을 때에는 정확성은 떨어지지만, 신뢰도, 즉 정밀도는 높다고 할 수 있다(그림 4-1-1).



그림 4-1-1. 타당도와 신뢰도의 개념

타당도는 결국 참값과 비교하여, 측정 자체가 정확한가에 대한 지표이다. 반면에 신뢰도는 반복 측정된 값 간의 비교로서 측정에 일관성이 있는가에 대한 지표이다. 그래서 타당도의 정의는 측정값이 측정하고자 목적하는 것을 반영하는 정도라 표현할 수 있고, 신뢰도는 여러 번 측정할 때 측정값이 같은 값에 접근하는 정도를 반영한다고 할 수 있다. 대표 통계량도 유추해 보면 타당도는 참값과 얼마만큼 떨어졌는가를 평가하면 되는 것이지만, 신뢰도는 반복 측정된 값의 평균과 표준편차를 이용한 분산을 평가하게 된다. 타당도 평가에는 민감도(sensitivity), 특이도(specificity), 위양성(false positive), 위음성(false negative), 양성 예측도(positive predictive value), 음성 예측도(negative predictive value) 등의 지표가 사용되며, 신뢰도 평가에는 일치(백분)율(agreement percent), 카파통계량(Kappa statistic), 상관 계수(correlation coefficient), 표준 오차(standard deviation), 변이 계수(coefficient of variation), 척도화 분석(Cronbach's alpha) 등의 지표 등이 있다. 여기서 측정이 여러번 정확하게 이루어졌다면, 즉 타당도가 일관성 있게 높다면 동시에 신뢰도도 높아지겠지만, 신뢰도가 높다고 해서 반드시 타당도가 높은 것은 아니다. 타당도를 높이기 위해서는 결국 계통오차(systemic error)를 줄여야 하는데, 이러한 계통오차는 일정한 방향으로 참값과 차이가 나는 것으로 바이어스(bias)라고 하기도 한다. 계통오차를 줄이기 위해서는 가장 쉬운 방법으로 측정 방법, 측정 도구를 표준화하여 정

도 관리를 하는 것이 있고, 실제 관찰자와 관찰대상의 상호 작용을 줄이기 위해 맹검법을 사용하기도 하고, 대상자의 기억에 의존하는 조사 대신에 다른 방법을 선택하는 것이다. 신뢰도를 높이기 위해서는 무작위 오차(random error)를 줄이는 것인데, 말 그대로 무작위로 발생하는 오차를 줄이기 위해서는 연구대상자 수, 즉 표본의 크기를 늘리는 방법이 가장 우선이고, 타당도에서와 마찬가지로 측정 방법 표준화, 조사자 표준 교육, 측정도구 자동화, 정밀성 등과 측정을 반복하는 것이다(표 4-1-1).

표 4-1-1. 타당도와 신뢰도의 비교

항목	타당도	신뢰도
정의	측정값이 측정하고자 목적하는 것을 반영하는 정도	여러 번 측정할 때 측정값이 같은 값에 접근하는 정도
핵심 질문	측정이 정확한가?	측정이 일관성이 있는가?
측정	참값과 비교	반복 측정한 값 간의 비교
대표 통계량	Mean difference[(참값-측정값)/참값]	분산계수(표준편차/평균)
연구 목적	결론의 타당도를 증대	효과를 측정해 내는 power를 증대
평가	민감도(sensitivity) 특이도(specificity) 위양성(false positive) 위음성(false negative) 양성 예측도(positive predictive value) 음성 예측도(negative predictive value)	일치율(agreement percent) 카파통계량(Kappa statistic) 상관 계수(correlation coefficient) 표준 오차(standard deviation) 변이 계수(coefficient of variation) 척도화 분석(Cronbach's alpha)
저해 요인	관찰자, 관찰대상, 측정도구에 의한 systemic error (bias)	관찰자, 관찰대상, 측정도구에 의한 random error (variance)

3 타당도의 평가

1. 타당도 지표

타당도는 민감도, 특이도, 위양성, 위음성, 양성예측도, 음성예측률 등으로 표현되는데, 모두 표 4-1-2에 있는 것처럼 수식으로 이해할 수 있다. 개념적으로는 민감도는 어떤 도구가 질병이 있는 사람을 얼마만큼 잘 찾아내는가를 나타내는 척도이며, 특이도는 질병이 없는 사람을 얼마큼 질병이 없다고 잘 분류하는가를 나타내는 척도이다. 민감도가 높아지려면 위음성이, 특이도가 높아지려면 위양성이 작아야

하는 것을 표에서 계산할 수 있다. 일반적으로 민감도가 높은 검사 방법은 중요한 질병을 찾아내는 것으로 위음성으로 치료 시기를 놓치지 않도록 하는 데 중요하고, 특이도가 높은 검사 방법은 질병이 없다는 것을 확인시켜 주어서 위양성으로 추가적인 검사나 걱정을 하지 않도록 하는 데 중점을 두고 있다. 예를 들어 질병의 유병률이 높은 질환의 경우에는 검사를 통해 질병이 아닌 것을 정확히 가려 주면 되기 때문에 특이도가 높아야 한다. 그렇지만 유병률이 낮고 위중도가 높은 암 같은 경우에는 민감도가 높아야 환자를 놓치지 않고 조기에 진단할 수 있게 된다.

표 4-1-2. 타당도의 평가

질병 유무 / 검사결과	있음(+)	없음(−)	합계
양성(+)	a (true positive)	b (false positive)	a+b
음성(−)	c (false negative)	d (true negative)	c+d
합계	a+c	b+d	a+b+c+d

- 민감도(sensitivity): $\frac{a}{a+c}$
- 특이도(specificity): $\frac{d}{b+d}$
- 위양성(false positive): $\frac{b}{b+d}$
- 위음성(false negative): $\frac{c}{a+c}$
- 정확도(accuracy): $\frac{a+d}{a+b+c+d}$
- 유병률(prevalence): $\frac{a+c}{a+b+c+d}$
- 양성 예측도(positive predictive value): $\frac{a}{a+b}$
- 음성 예측도(negative predictive value): $\frac{d}{c+d}$

2. 타당도 지표의 변화[1]

"민감도 90%로 허가받은 자가검사키트, 검사 현장에서 양성예측도가 76%인 이유는?"[1]

『(중략) 국내에서 사용하는 자가검사키트는 정확도가 90% 이상인 제품이라는 이야기, 뉴스에서 보신 적 있으실 텐데요. 하지만 앞서서 코로나19 진단 체계를 시행한 광주, 전남, 경기 평택, 경기 안성 4개 지역 선별진료소에서 자가검사키트로 양성 진단을 받은 사람 중 76.1%만 진짜 양성으로 나타나서 정확도가 76.1% 수준이라는 보도도 있었습니다.』

위 기사를 보면, 식품의약품안전처는 민감도 90%, 특이도 99% 이상으로 성능이 입증된 자가검사키트만 허가해 주었는데, 어떤 도구가 당연히 민감도와 특이도가 모두 100%에 가깝다면 정확도 면에서 더할 나위가 없을 것이다. 그런데 선별진료소에서 자가검사키트로 양성 진단을 받은 사람 중 76.1%만 진짜 양성으로 나타나 자가검사키트의 정확도가 그럼 어떻다는 것인지 의문이 생긴다. 76.1%를 찾아내는 정확도가 양성예측도인데 이러한 정확도가 높으면 당연히 좋다. 위의 수식에서 양성예측도는 민감도와 특이도가 고정되어 있다고 할 때 질병의 유병률에 영향을 많이 받게 된다. 이러한 결과는 또한 기기를 개발하는 환경과 실제 현장에서의 상황이 다르다는 점에서도 기인한다. 기기를 개발할 때는 성능 검사를 위해 적절한 비율로 질병 있는 군과 없는 군을 배정하여 검사하는 것이 일반적이다. 그렇지만 현실에서는 유병률의 개념으로 질병군과 비질병군의 분율이, 일반적인 질병일 경우 비질병군의 분율이 월등히 높아 실제 유병률은 낮다. 그런데 질병의 유병률이 높을수록 양성예측도는 같은 민감도와 특이도 하에서 높게 측정된다. 기기를 개발할 때는 정확성을 민감도와 특이도와 같은 지표를 활용하지만, 현장에서 활용할 때는 양성예측도과 같은 지표를 활용하며, 질병의 대유행 초기에 유병률이 낮을 때는 양성예측도가 낮아 검사를 해서 양성이 나오더라도 양성이 아닐 가능성(즉, 위양성)이 높은 것이다. 표 4-1-2에서 보더라도 유병률, 즉 전체 대상자 중 질병자의 분율이 높아지게 되면, 결국 a (true positve)가 커지는 것이며, a가 양성예측도를 결정하는 수식의 분자에 위치하고 있어서 결국 양성예측도가 증가하게 된다. 양성예측도를 수식으로 표현하면 아래와 같다. 또한, 유병률이 고정되어 있다고 가정할 때 민감도보다 특이도가 증가할 때 양성예측도에 더 영향을 주기도 한다.

1 식품의약품안전처. https://www.mfds.go.kr/brd/m_1056/view.do?seq=28&srchFr=&srchTo=&srchWord=&srchTp=&itm_seq_1=0&itm_seq_2=0&multi_itm_seq=0&company_cd=&company_nm=&page=2. 2022.02.16.

$$PPV=\frac{(\text{Prevalence}\times\text{Sensitivity})}{(\text{Prevalence}\times\text{Sensitivity})+(1-\text{Prevalence})\times(1-\text{Specificity})}$$

결국, 진단 검사법의 경우 민감도가 낮은 검사는 해당 질환의 발견이 어려우므로 조기진단의 기회를 놓치는 결과를 초래하게 되며, 특이도가 낮은 검사는 수검자에게 불필요한 걱정을 끼치며 힘들고 돈이 많이 드는 다음 단계의 검사를 유도하게 된다. 또 예측도가 낮은 검사는 검사결과의 해석에 혼동을 초래하므로 선별검사로 채택되는 검사는 높은 민감도와 특이도 및 높은 예측도를 가져야만 한다.

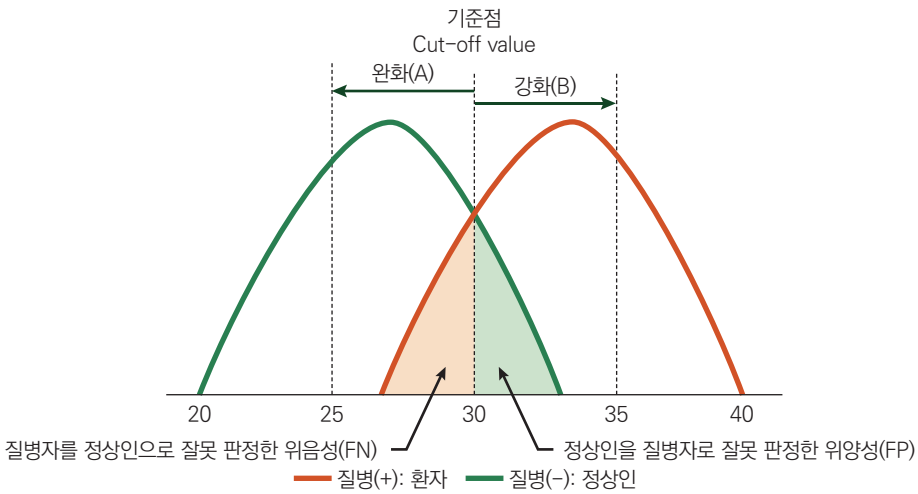


그림 4-1-2. 기준점 변화(완화, 강화)에 따른 민감도와 특이도의 변화

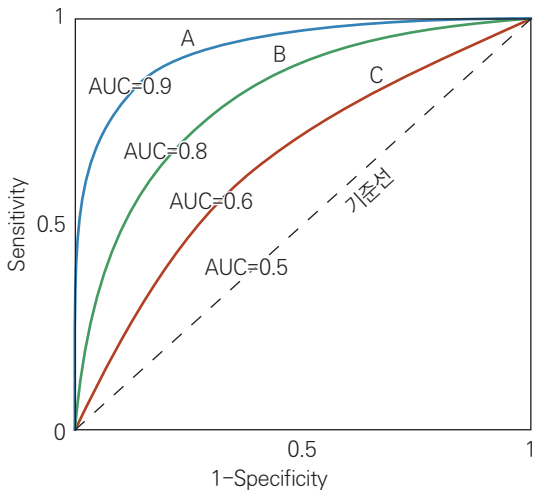
한편 민감도와 특이도는 사실 모두 100%가 된다면 이상적이겠지만, 현실에서는 그렇지 못하다. 같은 검사에서 민감도가 증가하면 특이도는 감소하고, 그 역도 성립하는 것을 관찰할 수 있다. 임상에서 우리는 많은 검사에서 기준점(cut-off value)을 사용하고 있다. 그림 4-1-2에서 검사 기준점을 30에서 25로 완화하게 되면, 질병에 걸린 사람(빨간색)은 모두 환자로 판정하여 민감도는 100%가 되지만, 질병이 없는 사람도 환자에 포함하게 되어 위양성이 증가하게 된다(특이도 감소). 반대로 기준점을 35로 강화하게 되면 질병이 없는 사람은 모두 제대로 판정하여 특이도는 100%가 되지만, 질병이 있는 사람을 환자로 진단하는 것을 놓쳐서 위음성이 증가하게 된다(민감도 감소).

사례, 코로나 검사

코로나19에서 RT-PCR 검사를 시행할 경우 Ct값(cycle threshold value)을 33.5를 기준으로 양성과 음성을 판정한 바 있다. 그럼, 여기서 RT-PCR 검사의 민감도와 특이도가 모두 100%인가에 대한 대답은 그렇지 않다는 것이다. 민감도와 특이도, 두 지표를 가장 만족하는 값이 33.5라는 것이다. 이때 Ct값을 32.5로 낮추게 되면 어떻게 될까? 32.5로 한다는 것은 기준을 좀 더 엄격하게 적용한다는 것으로 이렇게 되면 코로나19 환자가 아닌 경우는 더 음성으로 판정하지만(특이도 증가), 반대로 실제 코로나19 환자를 음성으로 판정할 수 있다(민감도 감소). 반대로 Ct값을 34.5로 증가시키면, 코로나19 환자는 확실히 가려주지만(민감도 증가), 상대적으로 코로나19 환자가 아닌 경우를 양성으로 판정하는 위양성이 증가하게 된다(특이도 감소).

3. ROC 곡선 분석

검사를 수행할 때나 검사 도구를 개발할 때 정확도가 가장 높은 검사 기기를 선택해야 하며, 그 기기가 어떤 기준점(cut-off value)으로 검사를 했을 때 가장 높은 정확도를 가질 수 있는가에 대한 고려가 필요하다(정상과 비정상을 잘 구분할 수 있는 능력). 무작정 어떤 기기의 민감도를 증가시킬 수는 없다. 위에서 말한 것처럼 그럼 특이도가 낮아질 수밖에 없는 것이다. 이런 경우 분석을 통해 통계적으로 가장 높은 민감도와 특이도를 가지는 기준점을 탐색하는 분석 방법이 ROC (Receiver Operating Characteristic) 곡선 분석 방법이다.



AUC값	해석
≥0.9	매우 좋음(Excellent)
0.8≤, <0.9	좋음(Considerable)
0.7≤, <0.8	보통(Fair)
0.6≤, <0.7	나쁨(Poor)
0.5≤, <0.6	의미 없음(Fail)

그림 4-1-3. ROC 곡선과 AUC

모든 진단검사의 특이도(specificity)와 민감도(sensitivity)는 선택된 기준점 값에 따라 달라진다. 따라서, 개별 기준점에 대해 하나의 민감도, 특이도 값 쌍이 존재한다. ROC 곡선은 이러한 값들을 연속해서 그래프에 나타낸 것이고, x축에는 1-특이도(1-specificity, 위양성), y축에는 민감도(sensitivity, 진양성)를 배치하여 구성된다. ROC 곡선에서 AUC(곡선 아래 영역, Area Under the Curve)는 검사법의 변별력을 평가하는 데 도움을 준다. 곡선이 왼쪽 위 모서리에 가까울수록, 그리고 AUC값이 클수록 그 검사가 질병이 있는 사람과 없는 사람을 더 잘 구별할 수 있음을 의미한다. AUC는 진단 정확도를 나타내는 전반적인 지표로 사용되며, 0과 1 사이의 값을 가진다. AUC값이 클수록 검사의 질이 우수함을 의미한다. 두 개의 ROC 곡선의 AUC를 비교하면 어느 검사가 질병 진단에 더 우수한지 평가할 수 있다. 그림 4-1-3에서 A검사>B검사>C검사 순으로 좋은 검사법인 것이다. 완벽한 진단검사는 AUC가 1.0이며, 0.5는 무작위(동전을 던져 앞, 뒤 어느 하나가 나올 확률은 0.5)로 분류한 것이므로 이 기준보다는 커야 하지만, 일반적으로 0.8 이상이 되어야 좋은 검사 도구로 수용이 가능하다.[2] AUC가 0.5보다 크더라도 이 부분도 통계적 검정(귀무가설, AUC=0.5)을 통해 유의성이 확보되어야 하는 것은 당연하다.

한편 기준점은 일반적인 통계 패키지에서 쉽게 계산되지만 가장 일반적으로 사용하는 방법은 유덴 지표(Youden index) (J)[3, 4]를 구하는 방법이다. 유덴 지표는 민감도와 특이도의 합에 1을 빼 준 결과로써 0과 1 사이의 값으로 계산되며, 값이 1에 가까울수록 잘 분류되었다고 볼 수 있다. 또한 AUC (Area Under the Curve)는 유덴 지표와 비례하는 값이므로, AUC가 최대가 되는 점을 찾는 방법과 같은 기준점을 가진다.

2 Sigismund AM. Diagnostic Accuracy? Part 1: Basic Concepts Sensitivity and Specificity, ROC Analysis, STARD Statement. Point of Care. The Journal of Near-Patient Testing & Technology 2012;11(1):6-8. https://doi.org/10.1097/POC.0b013e318246a5d6.

3 Youden, W. J. (1950). Index for rating diagnostic tests. Cancer 3, 32-35.

4 Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. Biom J 2005;47(4):458-72.

4 신뢰도의 평가

1. 양적 변수의 신뢰도 평가

연속변수로 이루어진 두 집단 간의 신뢰도는 상관관계, 즉 두 개의 양적 변수(quantitative variables) 간 방향성과 강도를 측정하게 된다. 상관관계를 보는 방법은 Pearson's r, Spearman's rho (rs), 그리고 Kendall's Tau (τ) 등이 있다. Pearson's r, Spearman's rho (rs)와 Kendall's Tau (τ) 등은 모두 상관관계를 나타내는 계수들이고, 단지 집단의 분포, 즉 연속변수로 구성된 집단이 정규 분포 여부에 따라 정규분포를 한다면 Pearson 상관계수를 이용하고, 그렇지 않다면 Spearman's rho (rs)와 Kendall's Tau (τ)을 이용하게 된다. 상관계수 r의 값은 -1에서 +1 사이의 범위를 가지며, r=0이면 두 변수 간에 아무런 상관관계가 없음을 의미하고[5], 일반적으로 0.6 이상이면 어느 정도 수용 가능한 좋은 상관관계로 판단할 수 있다.

- 0.1~0.2: 약한 상관관계(poor correlation)
- 0.3~0.5: 적당한 정도의 상관관계(fair correlation)
- 0.6~0.7: 중간 정도의 상관관계(moderate correlation)
- 0.8~0.9: 강한 상관관계(very strong correlation)
- r=1: 완전한 상관관계(perfect correlation)

2. 질적 변수의 신뢰도 평가

연속변수가 아닌 두 집단의 신뢰도는 일치(백분)율(agreement percent)과 카파통계량 또는 카파계수(Cohen's Kappa coefficient)[6]로 평가하는데, 예를 들어 두 명의 평가자가 특정 암을 대상으로 양성 여부를 평가하였을 때 두 명 모두 양성으로 판정한 경우와 두 명 모두 음성으로 판정한 경우가 두 명의 평가자가 일치하는 일치백분율이라 하고, 여기에서 두 명이 모두 우연히 양성 또는 음성으로 평가할 확률을 고려해서 일치도, 즉 코헨의 카파 계수를 산출할 수 있다. 산출식은 아래와 같다.

5 Chan YH. Biostatistics 104: correlational analysis. Singap Med J. 2003;44(12):614-619.

6 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159-74.

$$K=\frac{P(A)-P(B)}{1-P(B)}$$

여기에서 $P(A)$는 두 평가자 간 일치 확률(일치백분율)이고, $P(B)$는 두 평가자 간 우연히 일치할 확률이다. 표 4-1-3에서 $P(A)=\frac{a+d}{a+b+c+d}$로 표시할 수 있고, $P(B)$는 분리해서 평가자 1이 양성으로 판정할 확률은 $\frac{a+c}{a+b+c+d}$이고, 음성으로 판정할 확률은 $\frac{b+d}{a+b+c+d}$이다. 마찬가지로 평가자 2가 양성으로 판정할 확률은 $\frac{a+b}{a+b+c+d}$이고, 음성으로 판정할 확률은 $\frac{c+d}{a+b+c+d}$이다. $P(B)$는 두 평가자 모두 양성으로 판정할 확률과 모두 음성으로 판정할 확률을 더한 값이 된다. 수식은 아래와 같다.

$$P(B)=(\frac{a+c}{a+b+c+d}\times\frac{a+b}{a+b+c+d})+(\frac{b+d}{a+b+c+d}\times\frac{c+d}{a+b+c+d})=\frac{(a+c)(a+d)+(b+d)(c+d)}{(a+b+c+d)^2}$$

표 4-1-3. 일치도의 평가

판독		평가자 1		합계
		양성(+)	음성(−)	
평가자 2	양성(+)	a	b	a+b
	음성(−)	c	d	c+d
합계		a+c	b+d	a+b+c+d

카파 계수는 0.4 이상이면 중간 정도, 0.6 이상이면 강한 일치가 있는 것으로 판정할 수 있다.

표 4-1-4. Cohen의 카파 계수 해석

κ값 범위	해석(일치 수준)
≤0	일치 없음(Poor)
0.01≤, κ<0.20	매우 약한 일치(Slight)
0.20≤, κ<0.40	약한 일치(Fair)
0.40≤, κ<0.60	중간 정도의 일치(Moderate)
0.60≤, κ<0.80	강한 일치(Substantial)
0.80≤, κ≤1.00	거의 완전한 일치(Almost perfect)

참고 문헌

1. 식품의약품안전처[인터넷]. 기타 홍보자료 [Cited 2022 Feb 16]. Available from: https://www.mfds.go.kr/brd/m_1056/view.do?seq=28&srchFr=&srchTo=&srchWord=&srchTp=&itm_seq_1=0&itm_seq_2=0&multi_itm_seq=0&company_cd=&company_nm=&page=2.
2. Chan YH. Biostatistics 104: correlational analysis. Singapore Med J. 2003;44(12):614-9.
3. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. Biom J. 2005;47(4):458-72.
4. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.
5. Simundic AM. Diagnostic Accuracy? Part 1: Basic Concepts Sensitivity and Specificity, ROC Analysis, STARD Statement. Point of Care. The Journal of Near-Patient Testing & Technology. 2012;11(1):6-8.
6. YOUDEN WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32-5.

제2절 차세대염기서열 기반 유전체분석검사의 이해

분자생물학 기술의 급속한 발전으로 인해 유전체 분석 방법이 획기적으로 변화하고 있다. 특히 차세대염기서열분석(next generation sequencing, NGS) 기술의 등장으로 전장유전체 염기서열분석(whole genome sequencing, WGS)이 가능해졌으며, 이는 유전체학적 접근(genomic approach)의 새로운 지평을 열었다. 이러한 기술 발전은 전통적인 역학조사의 범위를 크게 확장시키며, 질병의 전파 경로를 추적하고 유전적 요인을 분석하는 데 있어 활용되고 있다.

WGS는 특정 유전자나 유전체의 일부가 아닌, 생물체 전체의 유전 정보를 분석하는 기술로, 역학조사에서 질병의 기원과 전파 경로를 밝히는 데 매우 유용하다. WGS는 세균 균주를 추적하는 데 있어 이전의 방법들로는 불가능했던 고해상도의 분석을 가능하게 한다. 예를 들어, 펄스필드겔전기영동(pulsed field gel electrophoresis, PFGE), 반복요소 PCR (repetitive DNA sequence-based PCR, rep-PCR), 다중위치 염기서열 타이핑(multi-locus sequence typing, MLST)과 같은 기존 방법들은 미생물의 핵산 일부만을 사용하여 균주를 비교하는 데 비해, 전장유전체 분석은 보다 정확하고 풍부한 정보를 제공한다. 이러한 특징을 이용하여 감염병 유행 시, 병원체의 유전체를 분석하여 병원체 간 유전적 유사성을 비교하고 전파 경로를 추적함으로써 감염 확산 방지를 위한 적절한 대응 방안을 수립할 수 있다. 또한, 병원체 유전체 분석은 항생제내성 유전자와 독소 유전자를 탐지하여 치료 전략을 최적화하고, 감염병의 추가 확산을 예방하는 데 기여한다. 팬데믹 상황에서는 특히 전장유전체 분석이 전파 속도를 분석하고 감염원을 신속히 확인하는 데 유용하다. 전장유전체 분석의 또 다른 중요한 응용은 신종 병원체의 탐지와 분류이다. 새로운 병원체가 출현할 경우, 유전체 분석을 통해 기존 병원체와의 유전적 연관성을 확인하고, 적합한 진단 방법과 치료법을 신속히 개발할 수 있다. 이는 감염병 초기 단계에서 효과적인 방역 조치를 취하는 데 있어 핵심적인 역할을 한다. 이번 절에서는 NGS를 기반으로 한 다양한 유전체 분석검사의 원리와 방법을 소개하여 실제 역학 조사 현장에서 검사 결과를 이해하는 데 도움이 되고자 한다.

1 차세대염기서열분석의 검사원리

1. 차세대염기서열분석의 개요

차세대염기서열분석(NGS)은 고전적인 DNA염기서열분석방법인 Sanger sequencing과 구별되는 새로운 염기서열분석 기술을 말하며, 다른 용어로는 massive parallel sequencing (MPS)이라고도 한다. NGS는 작은 조각으로 분해된 DNA가 합성될 때 염기마다 방출하는 신호를 분석하여 서열 데이터를 생성하는 방식으로 다량의 유전 정보를 얻기에 비용-효과적 측면에서 우수하여 기존 염기서열분석을 이용한 진단의 제한점을 극복할 수 있게 해주었다. NGS 검사의 단계는 크게 3가지로 구분되는데, ① 핵산의 전처리 과정, ② 장비에서 염기서열 생성(sequencing), 그리고 ③ 장비에서 산출된 데이터를 변환하여 분석하는 단계로 구성된다.

1) 핵산의 전처리 과정

(1) 분절화(Fragmentation)

핵산이 NGS 장비에서 합성 및 분석될 수 있도록 적절한 길이로 분절되는 과정을 말한다. 초음파 등의 물리적(sonication) 또는 효소적(nuclease, fragmentase, transposase 등) 방법으로 핵산을 무작위 절단하거나 엑손(exon) 등을 타깃으로 프라이머(primer)에 의한 증폭을 통해 핵산을 선택적으로 분절하게 된다.

(2) 라이브러리(Library)의 제작

① 말단수리(End repair)

분절화된 핵산에서 3′ 말단에 DNA Polymerase 등을 이용하여 상보적 염기로 채우거나 절단하고 5′ 말단에 polynudeotide kinase를 이용하여 인산기를 추가로 붙여준다.

② 3′-말단 아데닌 추가(A-tailing)

양쪽 말단이 수리된 DNA의 3-말단에 아데닌(adenine) 염기를 추가하는 과정으로 이는 연쇄체 형성(concatemerization)을 방지하고 티민(thymine) overhang을 가진 어댑터(adaptor)와 상보 결합이 용이하도록 한다.

(3) 어댑터의 부착(Adaptor ligation)

약 60~65 mer 크기의 염기서열을 가진 올리고뉴클레오티드(oligonucleotide)로서, NGS 장비가 인식할 수 있는 고유한 염기서열과 검체를 구별할 수 있는 고유한 염기서열인 인덱스(index) 혹은 바코드(barcode)를 포함하고 있다. 어댑터는 A-tailing까지 완료된 DNA와 ligase에 의해 쉽게 결합할 수 있도록 3′ 말단에 티민을 포함하고 있으며 보통 Y자 형태(forked) 혹은 U자 형태(hairpin)의 구조를 가지고 있다.

(4) 라이브러리의 증폭과 선별

어댑터와 분절된 핵산의 결합인 라이브러리는 시퀀싱을 하기에 양이 충분하지 않기 때문에 DNA polymerase를 사용하여 PCR 증폭 과정을 거친다. 라이브러리 증폭 과정 중에는 어댑터 간의 dimer 형성 등 예기치 않은 부산물이 생성되어 서열분석에 영향을 주기 때문에 자성 비드(magnetic bead) 또는 전기영동(예: Bioanalyzer)을 이용하여 이를 제거한다.

(5) 타깃의 선별(Target enrichment)

증폭된 라이브러리 중 검사 대상이 되는 타깃 핵산을 선별하는 과정을 말하며, PCR 프라이머로 증폭을 하는 앰플리콘(amplicon) 방식과 프로브(probe)를 이용하여 교합(hybridization)하는 캡쳐(capture) 방식으로 나뉜다. 타깃의 범위에 따라 다중유전자패널검사(targeted panel sequencing), 엑솜 시퀀싱(whole exome sequencing, WES), 전장유전체 염기서열분석(whole genome sequencing, WGS)으로 나뉠 수 있다. PCR 기반의 타깃 선별은 라이브러리 제작 시 프라이머를 이용하여 특정 확산 부위만 증폭하는 과정이 포함되므로 보통 라이브러리 제작과 함께 수행된다. 분석될 타깃의 사이즈, 핵산의 농도, 분석하고자 하는 변이의 유형 등을 고려하여 적절한 선별 기법을 적용할 수 있다

2) 장비에서 염기서열 생성(sequencing)

(1) 일루미나사의 cycle reversible termination

일루미나사(Illumina)의 기술은 DNA 가닥을 복제 및 합성하면서 염기서열을 분석하는 sequencing by synthesis 방식의 일종이다. 증폭되는 과정은 다음과 같다(그림 4-2-1). ① glass flow cell의 고정된 서열(support)에 DNA 라이브러리의 서열이 부착된다. ② 서열의 다른 한 끝쪽이 또 다른 support와 연결되어 bridge 형태가 형성된다. ③ DNA polymerase가 bridge 구조를 가진 서열의 primer를 인식하고, 상보적인 reverse 서열을 만든다. ④ 합성이 완료된 이중가닥(double stranded) DNA는 분리된다. ⑤ 결과적으로, forward와 reverse strand가 완성된다. ⑥ ②-③-④ 과정이 반복되

어 같은 DNA를 가진 forward, reverse 서열이 집단(cluster)을 이루게 된다. 생성된 bridge를 주형(template)으로 3차 가닥이 합성이 되는 과정이 반복되면서 n차의 무수한 가닥이 단일클론 DNA 집단(monoclonal DNA cluster)을 이루게 되는데 이 단일클론 DNA 집단에 형광이 부착된 상보적인 염기(modifed dNTP)가 삽입되며 광신호를 발생하고 세척되는 과정의 cycle을 반복하게 된다. 짧은 순간의 각 cycle마다 염기특이적인 광신호를 방출하게 되고 cluster 위치에서 순차적으로 발생하는 형광신호를 CCD (coupled-charge device) 카메라로 기록하여 장비가 염기서열 데이터를 저장하게 된다.

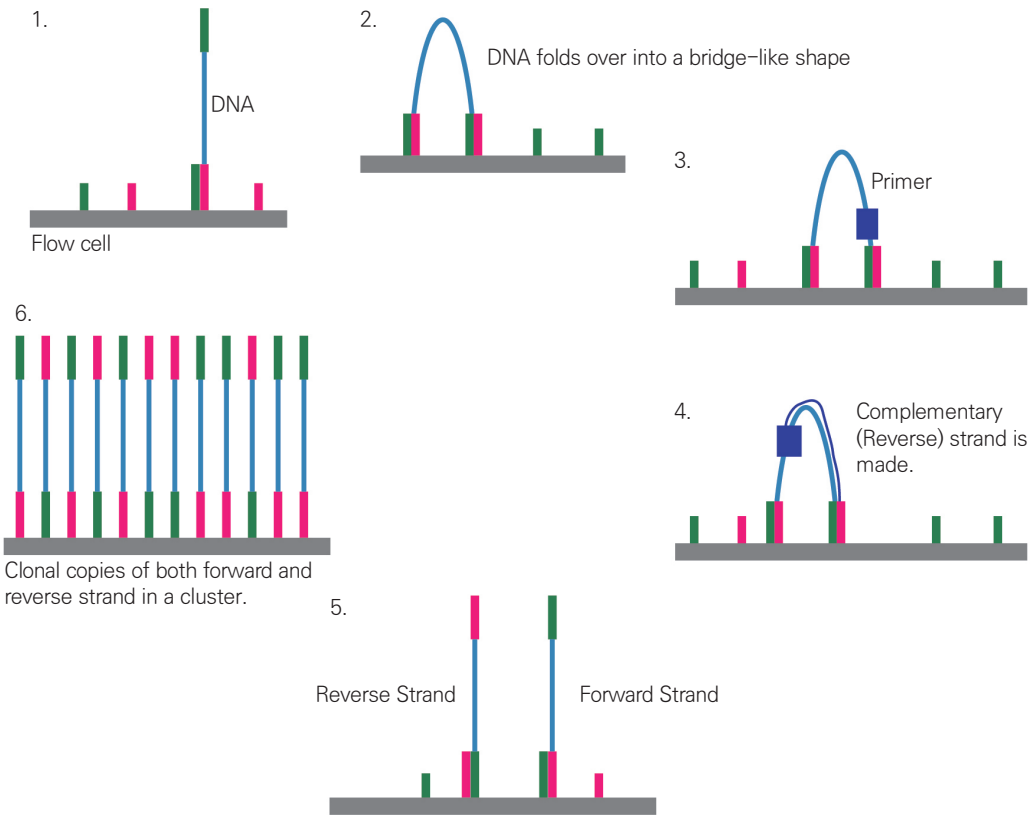


그림 4-2-1. 일루미나사의 sequencing by synthesis 과정
– Bridge amplification and clonal amplification

(2) 써모피셔사이언티픽사(ThermoFisher Scientific)의 semiconductor sequencing

Ion Torrent system의 기술로 라이브러리를 만든 후 작은 비드에 결합시키고 증폭한다. 라이브러리가 증폭된 후 well에 하나씩 들어가게 되면 well 바닥에 이온을 감지할 수 있는 트랜지스터가 전류를 측정하게 된다. 원리는 각 상보적인 dNTP (dATP, dTTP, dCTP, dGTP) 염기가 결합하는 합성과정에

서 방출되는 수소이온(H^+)에 의해 pH가 낮아지는 것을 측정한다(그림 4-2-2).

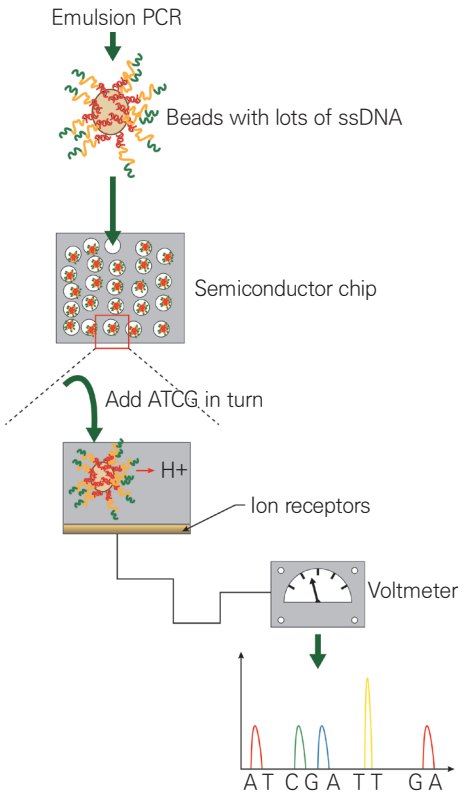


그림 4-2-2. 써모피셔사이언티픽사의 Semiconductor sequencing 과정

3) 데이터의 변환 및 분석

NGS 장비에서 생성된 시퀀싱 정보를 담은 파일은 바로 분석이 불가능하여 이를 변환시키는 과정이 필요한데 변환은 FASTQ → SAM → BAM → VCF 파일 순서로 변환과정을 거친다. FASTQ 파일은 짧은 단일 가닥의 염기서열 및 각 염기의 질 점수(quality score)가 기록된 텍스트 파일로, 유전체상의 위치 정보를 담고 있지 않다. 따라서, 참조 유전체(reference genome)에 정렬시켜 위치를 찾아주는 매핑(mapping)이 완료된 후에 각 시퀀싱 리드(read)별로 참조 유전체상의 위치가 기록된다. 이를 SAM (sequence alignment map) 파일이라 하며 이를 압축한 BAM (binary alignment map) 파일이 주로 분석에 사용된다. 유전체상의 위치 정보와 매핑의 정확도 등의 정보를 포함하는 SAM 또는 BAM 파일을 이용해 각 시퀀싱 리드의 서열이 참조 염기서열과 비교했을 때 관찰된 다른 변이(variation)가 있는지 찾아내는 변이의 검출(variant calling)을 거치며 이 과정을 통해 VCF (variant call format) 형

식의 파일이 생성된다. NGS 데이터 분석에서 FASTQ, SAM, BAM, VCF 파일과 같은 생물학적 데이터를 분석하고 시각화하는 데 사용되는 소프트웨어들을 생물정보학 도구(bioinformatics tools)라고 하며, 이러한 도구들을 순차적으로 사용하여 대량의 생물학적 데이터를 처리하고 분석하는 일련의 과정을 생물정보학 파이프라인(bioinformatics pipeline)이라고 한다.

2 역학조사를 위한 전장유전체분석

전장유전체분석(whole genome sequencing, WGS)은 앞서 설명한 NGS 과정을 통해 특정 병원체의 전체 유전체의 염기서열을 확보할 수 있는 기술로, 역학조사에서 병원체의 기원, 전파 경로, 유전적 다양성을 파악하는 데 매우 유용하다. WGS를 통한 역학조사는 다음과 같은 단계로 진행된다.

① 검체 수집 및 준비
② 유전체 시퀀싱(NGS)
③ 데이터 정렬 및 조립
④ 변이 분석 및 계통수 구축
⑤ 역학적 데이터 통합
⑥ 결과 해석 및 보고

1. 검체 수집 및 준비

WGS 분석의 첫 단계는 병원체가 포함된 검체를 수집하는 것이다. 검체는 감염 환자의 혈액, 조직, 타액, 객담 또는 환경적 샘플(예: 물, 음식 등)일 수 있으며, 역학조사의 목표와 분석 대상 병원체에 따라 적합한 샘플을 선택한다. 수집된 검체에서 병원체의 DNA 또는 RNA를 추출하여 NGS를 위한 준비를 한다. 이 과정에서는 샘플의 오염을 방지하기 위해 엄격한 무균 조건을 유지해야 한다. WGS 분석에 사용될 DNA는 NGS 장비에 투입되기 전에 라이브러리라는 형태로 준비된다. 라이브러리 제작은 크게 DNA 분절화, 말단 수선 및 어댑터 부착, 라이브러리 증폭 단계로 진행된다. DNA 분절화 과정에서 DNA 샘플은 일정한 길이의 조각으로 절단된다. 이를 위해 초음파 처리나 제한 효소를 사용하여 DNA를 분해한다. 절단된 DNA의 크기는 수백에서 수천 base pair (bp)에 이르게 된다. 말단 수선 및 어댑

터 부착에서는 절단된 DNA의 양 끝에 어댑터를 부착하여, 시퀀싱 기기가 이를 인식하고 읽을 수 있도록 한다. 어댑터는 시퀀싱 과정에서 필수적인 서열을 포함하고 있어, 각 DNA 조각의 위치를 정확히 추적할 수 있다. 라이브러리 증폭 단계에서는 어댑터가 부착된 DNA는 PCR 증폭을 거쳐 충분한 양을 확보한다. 증폭된 DNA는 품질 검사를 통해 NGS에 적합한지를 확인한다. 이러한 일련의 과정에서의 질 관리(quality control)는 시퀀싱 결과의 정확도를 높이는 데 매우 중요하다.

2. 유전체 시퀀싱

미생물의 유전체 시퀀싱(NGS)에는 3세대 염기서열분석법(third-generation sequencing)이 많이 활용된다. 3세대 염기서열분석법은 짧게 분절화 되지 않은 DNA 분자를 직접 타깃팅하여, NGS의 대량 병렬 특성을 유지하면서 1,000에서 수만 base pair에 이르는 long-read를 생성하여 'long-read NGS (LR-NGS)'라고 한다. 이와 구분하여 일반적으로 35~300 base pair의 read를 생성하는 경우를 'short-read NGS (SR-NGS)'라고 한다. 이러한 LR-NGS 기술은 3세대 염기서열분석법으로 불리며, Pacific Biosciences사의 single-molecule real-time (SMRT)와 Oxford Nanopore사의 플랫폼이 이에 해당된다. LR-NGS는 특히 플라스미드(plasmid) 시퀀싱에 유용하다. 플라스미드는 반복 영역(repetitive region)과 이동성 유전 요소(mobile genetic element)를 포함하고 있어 short-read 시퀀스 데이터로는 데이터 분석과정 중 assembly를 수행하기 어렵다. 반면 long-read 시퀀스 데이터는 플라스미드 염기서열의 효율적인 분석이 가능케 하여 동일 또는 다른 종의 분리주 간에 수평적 플라스미드 전이가 발생했는지 여부를 확인하는 데 사용될 수 있다.

1) Pacific Biosciences사의 Single Molecule Real-Time (SMRT)

Pacific Biosciences사의 SMRT 기술은 수백만 개의 개별 긴 DNA 단편을 읽어 평균 약 10,000 base pair read를 생성한다. 이 기술에서 사용되는 헤어핀 어댑터는 긴 DNA 분자를 여러 번 반복해서 읽을 수 있게 하여 SR-NGS에서와 같은 높은 정확도의 결과를 만들 수 있다. 그러나 높은 비용과 SR-NGS에 비해 낮은 처리량이 SMRT 기술의 더 광범위한 활용을 제한하는 주요 요인이다. 하지만 이러한 단점들에도 불구하고 LR-NGS는 SR-NGS와 비슷한 수준의 정확도를 가지면서 길게 읽을 수가 있어 특히 세균과 바이러스와 같은 비교적 작은 게놈 분석에 적합하여 미생물의 전장유전체 염기서열분석에 활용되고 있다.

2) Oxford사의 Nanopore

Oxford사의 Nanopore 기술은 DNA 분자가 장치의 막에 포함된 나노포어(nanopore)를 통과할 때 발생하는 뉴클레오타이드 특이 이온 전류 변화를 감지하는 방식으로 작동한다. 이 방법은 SMRT보다 낮은 정확도가 단점이다. 하지만 2014년에 등장한 이 기술은 실시간으로 데이터를 생성할 수 있고 휴대용 장치가 가능하다는 이점으로 병원체 감시 및 임상 진단에서 점점 더 많이 활용되고 있다. MinION 장치인 경우는 매우 소형으로 길이가 3인치에 불과하며, 노트북의 USB 포트로 전원이 공급되어 원격지나 자원이 제한된 환경에서도 신속히 배치할 수 있다.

이러한 제3세대 접근 방식은 아직 높은 처리량과 정밀한 정확도가 부족하여 단독으로 SR-NGS 기술을 완전히 대체하기 어렵다. 그러나 미생물은 데이터 정렬(alignment)과 조립(assembly) 과정에 필요한 참조 유전체(reference genome) 염기서열이 없는 경우가 많기 때문에 시퀀싱 리드를 조합하여 균의 염기서열을 추정하는 새로운 조립(de novo assembly) 알고리즘을 사용한다. 이때 LR-NGS는 SR-NGS보다 염기서열 데이터의 assembly 과정이 용이하여 미생물 유전체의 반복 영역 변이를 더 잘 평가할 수 있다. 또한, SR-NGS 장비의 short-read 데이터를 3세대 염기서열분석 장비에 추출한 long-read에 정렬하여 오류를 교정하는 하이브리드 시퀀싱(hybrid sequencing)이 가능하다.

3. 데이터 정렬 및 조립

WGS에서 얻어진 원시데이터(raw data)는 여러 분석 단계를 거쳐 유의미한 결과로 가공된다. 주요 분석 과정은 정렬(alignment), 조립(assembly), 주석(annotation) 단계로 나눠진다. 정렬 단계에서는 시퀀싱된 리드를 참조 유전체에 정렬하여 병원체의 유전체를 분석한다. 주요 도구로는 BWA, Bowtie, SOAP2 등이 있으며, 정렬 품질은 mapping quality와 coverage depth로 평가된다. 조립 단계에서는 정렬된 데이터를 바탕으로 유전체를 재구성하는 과정이다. 참조 기반 조립은 이미 알려진 유전체를 기반으로 하며, De novo 조립은 참조 유전체 없이 시퀀싱된 데이터를 통해 유전체를 재구성한다. N50과 contig의 수는 조립의 질관리에 사용되는 주요 지표들이다. 주석 과정은 조립된 유전체 서열에 대해 유전자나 기능적 영역을 식별하고 예측하는 과정이다. 이를 통해 유전체 내 각 영역이 어떤 기능을 수행하는지 파악할 수 있다.

4. 변이 분석 및 계통수 구축

정렬 및 조립된 데이터를 기반으로 병원체의 유전적 변이를 분석한다. 변이 분석을 통해 병원체 간의

관계를 파악하고, 이를 바탕으로 계통수(phylogenetic tree)를 구축하여 병원체의 기원과 진화 경로를 추적한다.

5. 역학적 데이터 통합

WGS 데이터를 다른 역학적 데이터와 통합하여 감염병의 전파 경로, 확산 양상, 감염원 추적 등을 분석한다. 이 통합 분석은 공공 보건과 질병 관리에 중요한 정보를 제공한다.

6. 결과 해석 및 활용

마지막으로 분석된 결과를 바탕으로 역학조사관, 임상의, 연구자들이 결과를 해석하고 보고한다. WGS의 결과는 다양한 역학적 질문에 답하는 데 활용되며, 병원체의 동정 및 특성 파악, 항생제내성 유전자 확인, 독성 인자 및 병원성 유전자 파악, 전파 경로 추적, 진화 연구 등의 분야에 활용될 수 있다.

3 메타게놈 분석

미생물을 동정하기 위한 'gold standard' 검사방법으로 배양법을 사용하고 있으나 실제로 검사실과 같은 인위적인 환경에서 배양이 가능한 미생물은 매우 극소수에 불과하고 또한 미생물의 배양에는 수일 정도의 시간이 많이 소모되는 단점이 있다. 최근 NGS 분석기술의 발전으로 배양을 하지 않고 검체 내 미생물의 전장유전체 염기서열분석이 가능한 metagenomics 분야가 비약적으로 발전하고 있다. Metagenomics는 "Meta"와 "Genomics" 두 단어로 이루어져 있다. Meta는 많은 유기체를 의미하고 Genomics는 유전자를 총체적으로 연구하는 학문을 의미한다. 즉 metagenomics는 다양한 환경(공기, 물, 토양, 인체 등)에 존재하는 미생물 전체의 genome을 한꺼번에 유기적으로 연구하는 유전체학의 한 분야를 말한다.

감염병의 역학조사에서 메타게놈 분석(metagenomic analysis)은 중요한 역할을 한다. Metagenomics 분석은 배양과정 없이 검체 내에 존재하는 모든 미생물의 유전 정보를 추출하여 분석하므로, 배양이 어려운 미생물이나 미지의 병원체를 신속하게 파악할 수 있는 장점이 있다. 이러한 특징으로 metagenomics은 기존의 배양법이나 PCR 기반 검사로는 한계가 있었던 다양한 병원체를 동

시 다발적으로 분석할 수 있는 강력한 도구이다. Metagenomics 분석방법에는 모든 세균에 공통적으로 존재하는 16S rRNA 유전자를 분석하는 방법과 검체에 존재하는 모든 유전체를 분석하는 shotgun metagenomic sequencing 방법이 많이 사용된다.

1. 16S rRNA 유전자 분석

16S rRNA 유전자는 세균에서 공통적으로 존재하는데 약 1,500 bp 정도의 크기이며, 여기에는 균종마다 다르고 9개의 다양성(variability)이 큰 부위(variable region)가 존재한다. 이러한 변동성 부위 사이에는 균종에 상관없이 염기서열이 동일한 부위(constant region)가 존재하는데, 이 부위에 맞추어 모든 균이 공통적으로 증폭될 수 있는 유니버설 프라이머(universal primer)를 고안하여 PCR 증폭을 하여 미생물의 속(genus) 또는 종(species)과 같은 계통발생학적 분류를 하는 데 활용될 수 있다. 이 분석 방법은 16S rRNA 유전자 부위만을 증폭하는 NGS 방법을 사용하기에 amplicon-based next generation sequencing 방법이라고 불리며 다른 유형으로는 곰팡이에서 발견되는 ITS 영역을, 원생생물은 18S rRNA 영역을 증폭하여 분석이 가능하다. 이 방법은 특히 대규모의 미생물 군집을 분석하는 데 매우 효과적이다. 16S rRNA 분석을 통해 미생물 군집의 분포를 파악하고, 각 균종의 상대적 비율을 추정할 수 있다. 이 분석은 비교적 경제적이며, 빠르고 정량적인 결과를 얻을 수 있다.

2. 샷건 메타게놈 시퀀싱

샷건 메타게놈 시퀀싱(shotgun metagenomic sequencing) 방법은 샘플에서 추출한 전체 DNA를 무작위로 절단하여, 각 조각을 시퀀싱한 후, 이를 이어 붙여 샘플 내에 존재하는 모든 미생물의 유전체 정보를 분석한다. 샷건 메타게놈 시퀀싱 방법은 16S rRNA 유전자 분석보다 더 깊이 있는 정보를 제공할 수 있다. 전체 유전체를 분석할 수 있기 때문에, 세균뿐만 아니라 바이러스, 곰팡이, 플라스미드, 박테리오파지 등의 다양한 미생물을 동시에 식별할 수 있다. 또한, 특정 미생물의 유전자 변이나 병원성 특성을 파악할 수 있어 역학적 연구에 매우 유용하다. 하지만, 샷건 메타게놈 시퀀싱 방법은 상대적으로 높은 비용과 시간이 소요되므로, 필요한 경우에만 사용된다.

샷건 메타게놈 시퀀싱의 장점은 전체 유전체를 분석할 수 있다는 것이다. 이를 통해 단순히 종(species) 정보를 넘어서, 균주의 세부 정보(stain)까지 파악할 수 있으며, 미생물의 유전자 기능이나 내성 관련 돌연변이까지 식별할 수 있다. 그러나 이 방법은 host DNA를 제거하여 분석해야 하며, 이를 통해 불필요한 분석 비용을 줄일 수 있다. 반면 16S rRNA 유전자 분석은 상대적으로 빠르고 경제적인 방법

으로, 특정 미생물 군의 분류를 정확히 수행할 수 있으며 알려진 reference 정보를 통해 미생물의 기능을 예측할 수 있다.

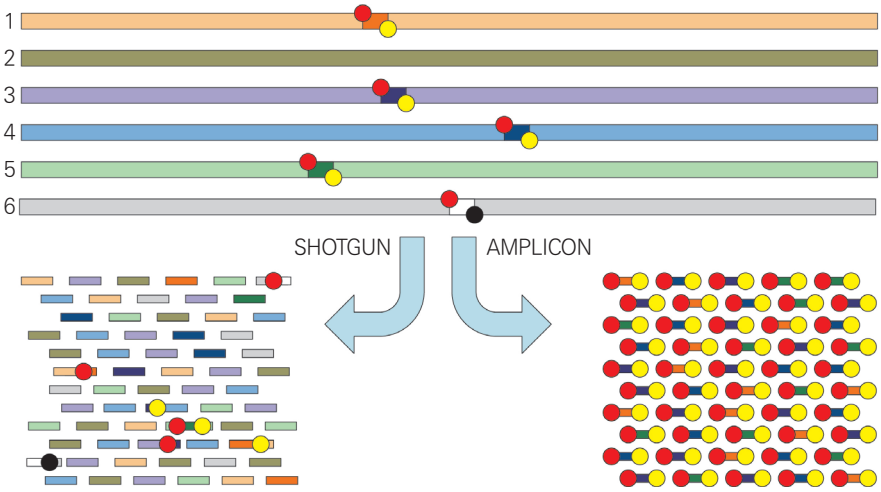


그림 4-2-3. Shotgun metagenomic sequencing과 amplicon-based sequencing의 차이를 나타낸 모식도

3. 역학조사에서 메타게놈 분석의 활용

Metagenomics 분석은 감염병의 역학조사에 있어 중요한 역할을 한다. 예를 들어, 특정 감염병의 원인균을 정확히 식별하고, 그들의 유전자 변이를 분석하여 발병의 원인과 전파 경로를 추적할 수 있다. 또한, metagenomics 분석을 통해 감염자의 임상정보와 병원체의 유전적 특성을 결합하여 더 정교한 역학 모델을 개발할 수 있다. 이를 통해 질병의 확산을 예방하고, 효과적인 치료법을 개발할 수 있는 기초자료를 제공할 수 있다. Metagenomics 분석에서 얻은 데이터는 미생물 군집의 다양성을 분석하는 데도 유용하다. 환경에서 채취한 표본을 분석하여, 특정 환경에서 발견되는 미생물의 분포와 그들의 상호작용을 파악할 수 있다. 이러한 연구는 환경오염이나 새로운 병원체의 출현을 감지하는 데 중요한 역할을 한다. 또한, metagenomics 분석을 통해 미생물의 잠재적 기능을 연구할 수 있으며, 항생제내성 또는 바이러스의 변이에 대한 연구에도 활용될 수 있다. Metagenomics 분석은 감염의 역학조사에서 중요한 도구로 자리 잡고 있으며, 다양한 병원체를 신속하고 정확하게 식별할 수 있는 가능성을 제공한다. 배양법의 한계를 극복하고, 기존의 검사법으로는 파악할 수 없는 미생물을 검출할 수 있는 metagenomics 분석은 앞으로도 감염병 연구와 예방, 치료에 중요한 기여를 할 것으로 예상된다.

4 전장유전체 염기서열분석 데이터를 활용한 역학조사 도구

1. 핵심 유전체 다중 유전자좌 서열 분석

핵심 유전체 다중 유전자좌 서열 분석(core genome multilocus sequence typing, cgMLST)은 세균의 핵심 유전자(core gene)를 기반으로 한 유전체 분석 기법으로, 전장유전체 데이터를 활용하여 세균의 계통 및 역학적 관계를 고해상도로 파악할 수 있는 도구이다. cgMLST는 기존 multi-locus sequence typing (MLST)의 확장된 형태로, 몇 개의 하우스키핑(house keeping) 유전자를 분석하던 MLST와 달리 핵심 유전자 전체를 대상으로 한다는 점에서 차별화된다. 이를 통해 보다 높은 분해능을 제공하며, 특히 동일 종 내의 세부적인 균주 차이를 구별하는 데 유용하다. cgMLST는 전장유전체 시퀀싱(whole genome sequencing, WGS)을 기반으로 하여 특정 종의 핵심 유전자 세트를 정의하고, 이들 유전자의 대립형질(allele) 정보를 비교하여 균주 간의 유전적 차이를 분석한다. 핵심 유전자는 해당 종의 대부분의 균주에서 보존되며, cgMLST 스키마(scheme)는 이러한 유전자들의 집합으로 구성된다. 예를 들어, *Listeria monocytogenes*에 대한 cgMLST 스키마는 1,701개의 핵심 유전자를 포함하며, 이들 유전자는 99% 이상의 균주에서 공통적으로 발견된다.

cgMLST는 고해상도 역학조사 도구로서 병원체의 계통학적 분석과 감염병 관리에 중요한 역할을 한다. 이 방법은 기존 MLST보다 더 많은 유전자 좌위를 포함하여 균주 간 미세한 차이를 구별할 수 있어 고해상도 분류를 가능하게 한다. 또한, cgMLST는 데이터 분석 및 결과 공유를 위한 표준화된 명명법을 제공하여 실험실 간 데이터 비교와 협력을 용이하게 한다. 이를 통해 병원체의 전파 경로를 추적하고, 지역 및 전 세계적 수준에서 발생 원인을 규명하는 데 효과적인 도구로 활용될 수 있다.

2. 전장유전체 단일염기다형성 분석법

전장유전체 단일염기다형성 분석법(whole genome single nucleotide polymorphism analysis, wgSNP)은 전체 유전체에서 단일 염기 다형성(SNP)을 식별하여 병원체의 유전적 변이를 분석하는 방법이다. 이 방법은 전장유전체분석데이터(WGS데이터)를 활용하여 병원체의 유전적 차이를 매우 세밀하게 분석할 수 있다. wgSNP 분석은 단일 염기 수준의 차이를 검출하여, 감염병 역학조사에서 유전적 변이 패턴을 정확히 파악할 수 있는 강력한 도구로 활용되고 있다. wgSNP 분석법의 가장 큰 장점은 높은 해상도와 포괄성이다. 기존의 유전형 검사법에 비해 높은 해상도로 병원체의 전체 유전체(예: 결핵

균의 경우 약 4.4 Mb)의 90% 이상을 분석할 수 있으며, 이로 인해 세밀한 유전자 변이를 검출할 수 있다. 또한, 전체 유전체를 분석함으로써 유전적 변이의 전반적인 패턴을 파악할 수 있어, 감염 경로 및 전파 양상을 추적하는 데 매우 유용하다. wgSNP 분석법은 다양한 병원체에 적용 가능하며, 새로운 변이를 발견하는 데도 유리하다.

wgSNP 분석의 과정은 전체 유전체 시퀀싱, 데이터 전처리, 참조 유전체 매핑, SNP 호출, SNP 필터링, 그리고 계통 분석을 포함한다. 먼저 고품질의 DNA를 추출하여 차세대 시퀀싱 기술을 이용해 전체 유전체 서열을 확보하고, 시퀀싱 데이터를 필터링하여 품질을 확인한다. 이후 얻어진 데이터를 참조 유전체에 매핑하여 SNP를 식별하고, 신뢰할 수 있는 SNP만을 선별하여 계통수를 작성하고 유전적 관계를 분석한다. wgSNP 분석법에는 몇 가지 한계점과 고려사항이 존재한다. 첫째, 대량의 유전체 데이터를 처리하고 분석하는 데 높은 컴퓨팅 능력과 전문 지식이 요구된다. 둘째, 분석 방법과 기준의 표준화가 필요하다. 이는 여러 연구 간의 결과 비교를 용이하게 하고, 신뢰성을 높이는 데 중요하다. 셋째, 적절한 참조 유전체의 선택이 분석 결과에 큰 영향을 미칠 수 있다. 마지막으로, 재조합이 빈번히 일어나는 영역에서는 SNP 분석이 어려울 수 있다.

wgSNP 분석은 다양한 감염병의 역학조사에 성공적으로 적용되고 있다. 특히 결핵, *Salmonella*, *E. coli*, *Campylobacter* 등에서 병원체 간 집단 발생 사례 분석에 유용하다. 예를 들어, 2015년부터 2018년까지 국내 4개 대학교에서 발생한 결핵 집단 발생 사례에서는, 기존의 spoligotyping 및 24 유전자좌 mycobacterial interspersed repetitive unit-variable number tandem repeat (MIRU-VNTR) 검사법으로는 감염 경로를 확인할 수 없었으나, wgSNP 분석을 통해 동일 감염원에 의한 집단 발생임을 재확인할 수 있었다. 반면, 대학 간 집단 간 유전자 변이가 큰 차이를 보였고, 이를 통해 3개 대학교에서 발생한 집단 발생 사례는 서로 연관성이 없음을 확인하였다.

wgSNP 분석법은 cgMLST와 다르게 비코딩 영역을 포함한 세균 유전체의 단일 뉴클레오티드 변이 호출을 기반으로 하여 분석에 더 높은 수준의 해상도를 추가할 수 있다. 이러한 wgSNP 분석법은 감염병 역학조사에서 매우 유용한 도구로, 기존의 유전형 검사법보다 높은 해상도와 정확성을 제공한다. 이 방법은 복잡한 감염 경로와 전파 양상을 파악하는 데 큰 도움이 되며, 앞으로 기술의 발전과 표준화가 이루어짐에 따라 그 활용도는 더욱 높아질 것이다.

5 기타 역학조사를 위한 검사방법들

공공 보건과 의료 역학 전문가들이 분리된 균주의 관계를 규명하려고 시도한 이후, 수십 년 동안 다양한 타이핑 방법들이 사용되었다. 주로 사용된 방법들로는 펄스필드겔전기영동(PFGE), 반복요소 PCR (rep-PCR), 다중위치 염기서열 타이핑(MLST) 등이다.

1. PFGE

PFGE 방법에서는 세균 DNA가 제한 효소에 의해 절단되고, 그 결과로 생성된 큰 DNA 조각들이 다방향 전기 펄스를 이용해 크기에 따라 분리된다. 각 균주의 결과 밴드를 비교하고, 설정된 기준을 사용해 균주가 관련이 있는지 없는지를 판단한다. 이 방법은 균주들 간에 밴드 패턴이 충분히 다르면 유용하지만, 유사하거나 동일한 밴드 패턴은 관계를 대략적이고 질적으로만 추정할 수 있다. 또한 이 기술은 시간이 많이 걸리고(최소 3일 이상), 작업이 번거롭고 처리량이 적다는 제한점이 있다.

2. rep-PCR

rep-PCR은 유전자 코드 사이에 위치한 수많은 반복적인 DNA 서열의 배열을 사용하여 세균 균주를 구분한다. 이 과정은 자동화할 수 있으며, 반복적인 서열을 증폭시켜 다양한 길이의 증폭산물을 생성하고 이를 전기영동으로 분리한 후, 소프트웨어를 이용해 균주 간 밴드 패턴을 읽고 비교할 수 있다. 이러한 패턴을 이용하여 세균 균주 고유의 genomic fingerprinting을 작성할 수 있다. PFGE와 마찬가지로 이 기술은 균주가 다른 균주에 속한다는 것을 보여줄 때 가장 유용하지만, 유사한 균주일지라도 여전히 반복서열에는 차이를 보일 수 있다는 단점이 있다.

3. MLST

MLST는 대략 7개 정도의 housekeeping 유전자의 400~500 bp 염기서열을 비교한다. 이 서열은 특정 프라이머를 사용해 증폭된 후, 증폭된 DNA 조각을 시퀀싱하여 각 위치에서 염기서열을 분석한다. 각 서열 조합은 균주의 염기서열 타입을 정의하는데, 이 분류는 전 세계적으로 표준화되어 있다. 전장 유전체분석이 가능해지면서, MLST는 해당 세균 균주의 전체 유전체 시퀀싱에서 식별된 동일한 위치와

염기서열을 사용하여 컴퓨터 시뮬레이션인 *in silico* 방식으로 구현이 가능하게 되었다. 또한, 전장유전체분석에서는 세균의 전체유전체의 염기서열분석이 가능하여, MLST 방법에서 일반적으로 사용되는 7개의 housekeeping 유전자에 제한되지 않고 더 많은 염기서열 타이핑 생성이 가능하다. 예를 들어, 핵심 유전자(core genes)로 자주 사용되는 수천 개의 위치를 이용한 전장유전체분석 기반 MLST 방법은 MLST의 해상도를 확장시켰다. 이 기술은 핵심유전체 MLST (core genome MLST, cgMLST)라고 불리며, 비교에 사용되지 않는 위치에서 발생할 수 있는 염기서열 변이를 놓칠 수 있지만, 전체 유전체 비교에 비해 계산 작업과 전문 지식이 덜 필요하고, 다른 연구기관에서도 적용 가능한 분류 체계를 생성할 수 있는 장점이 있다. 최근에는 전장유전체분석의 가격이 저렴해지면서, 정통적인 MLST 방법의 비용적인 메리트가 상대적으로 떨어지게 되었다.

참고 문헌

1. 식품의약품안전처. 차세대염기서열분석(Next Generation Sequencing) 검사의 임상적 변이해석 및 검증방법 해설서. 의료기기안전국; 2022.
2. 질병관리청 감염병진단분석국 세균분석과, 김동혁, 송승은, 전정훈, 하지민, 황규잠. 전장유전체 염기서열분석 유전형검사법(wgSNP)을 통한 결핵 집단발생의 감염경로 분석. 주간 건강과 질병. 2021.
3. Chiang AD, Palmore TN. Whole Genome Sequencing for Outbreak Investigation. In: Bearman G, Hota S, Morgan DJ, et al. Infection Prevention: New Perspectives and Controversies. 2nd ed. New York: Springer; 2022. pp. 223–36.
4. Budimir A, Laurent F, Nashev D, et al. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill. 2013;18:20380.
5. Dobrindt U, Holst-Jensen A, Johannessen GS, et al. High throughput sequencing for detection of foodborne pathogens. Front Microbiol. 2017;8:2029.
6. Gilchrist CA, Hewlett EL, Petri WA Jr, et al. Whole-genome sequencing in outbreak analysis. Clin Microbiol Rev. 2015;28:541–63.
7. Marchaim D, Nutman A. How to: molecular investigation of a hospital outbreak. Clin Microbiol Infect. 2019;25:688–95.
8. Pallen MJ, Robinson ER, Walker TM. Genomics and outbreak investigation: from sequence to consequence. Genome Med. 2013;5:36.

Textbook For
Korea Field Epidemiology
Training Program





5장

예방접종관리 역량강화







제1절 관찰역학연구를 통한 지역사회 백신효과의 측정
제2절 예방접종 의향에 대한 이해와 소통 중심의 대응 방안

집필자 유명순, 천병철

제1절

관찰역학연구를 통한 지역사회 백신효과의 측정

1 백신의 효과 측정

1. 백신의 효과와 효능

백신이 질병(혹은 감염)을 예방하는 능력을 측정하는 지표에는 효과(effectiveness)와 효능(efficacy)이 있다(영어의 'effectiveness'와 'efficacy'는 매우 다른 개념이어서 이를 모두 우리나라 말로 '효과'라고 같은 용어로 번역하면 혼란이 생긴다. 최근에는 efficacy를 효능이라고 달리 번역하여 혼란을 줄이고 있고, 여기서도 effectiveness는 효과, efficacy는 효능이라고 정의하였다). 백신 효능(efficacy)은 무작위 임상시험처럼 외부의 요인이 적절하게 통제된 이상상태에서 측정된 백신의 예방능력이다. 통상 백신의 임상시험에서도 무작위배정을 통하여 백신군과 비백신군의 비교성을 높이고, 관찰방법과 관찰기간, 질병발생여부를 모두 통제하여 관찰하기 때문에 이상적인 상황에서의 백신의 질병 예방능력이라고 말할 수 있다. 또한 효능은 대부분 백신이 시판허가를 받기 전에 측정된다. 일단 백신이 시판되면 실제 접종인구는 지역사회에 거주하는 인구가 된다. 지역사회인구는 임상시험에 참여한 사람과 성, 연령은 물론 기저질환 여부, 사회 및 가족의 접촉여부 등이 다르며, 지역사회마다 여러 특성이 다른 현실세계(real-world)이다. 이와 같은 현실세계에서 관찰된 백신의 질병예방능력을 백신 효과(effectiveness)라고 하며 임상시험에서 측정된 예방능력(효능)보다 통상 낮게 측정된다. 그러나 백신시판 허가 후 실제 백신을 사용하는 곳은 지역사회이며 따라서 공중보건학적으로 중요한 것은 임상시험에서 측정된 백신 효능보다는 지역사회에서의 백신효과이다. 즉, 임상시험에서 제외되는 인구집단(소아, 임산부, 심각한 질병의 위험이 높은 사람 등)도 지역사회에서는 포함되어 있으며, 지역사회에서의 접종은 백신의 보관 및 취급이나 백신 스케쥴의 순응 정도 등이 임상시험처럼 엄밀히 통제되는 것이 어려우며 실제 유행하고 진화하는 바이러스(예: 독감, COVID-19)들이 백신 성능에 계속 영향을 미치기 때문에 지역사회에서 백신의 효과를 측정해야 한다. 또 지역사회에서 백신의 효과는 백신을 맞은 사람에게만 나타나는 것이 아니라 집단면역에 의해서 백신을 안 맞은 사람에게도 나타난다. 이러한 이유 때문에 많은 나라에서 시판

후 백신의 실제 효과를 평가하고 모니터링하는 체계를 갖추고 있는 이유이다.

백신의 효과에 대한 용어중에 백신의 영향(impact)이 있는데 이것은 백신이 인구집단에 가져온 효과를 통칭하는 것이다. 영향에는 효과뿐 아니라 해당 인구집단의 백신접종률, 이상반응 정도와 이에 대한 태도 등 다양한 요인이 연관되어 있다.

2. 백신 효과의 정의

백신의 효과는 백신을 맞은 사람이 안 맞은 사람에 비해서 질병에 덜 걸릴 확률로 정의되며, 통상 %로 나타낸다. 즉, 백신을 맞은 사람과 안 맞은 사람의 발생률(rate) 또는 발생위험(risk)을 비교하는 것이며 수식으로는 다음과 같이 표현된다.

$$\text{식 1: } VF=\frac{(R_u-R_v)}{R_v}=1-\frac{R_v}{R_u}=1-RR$$

여기서 VE는 백신 효과, R_u와 R_v는 각각 비백신군과 백신군에서의 발생률(혹은 발병률 또는 발생위험)이고 RR은 상대위험도(relative risk)이다. 즉, 백신의 효과는 수학적으로 (1-상대위험도)로 정의된다. 이것은 코호트연구로 백신군과 비백신군의 발생률을 알 때 가능한 수식이다. 만약, 백신의 효과를 측정하기 위한 사례-대조군연구(case-control study)를 수행했다면 여기서 측정한 오즈비(odds ratio, OR)가 RR과 유사해지는 조건(해당 질환의 발병률이 10% 미만으로 낮아야 한다)을 만족하는 경우, 사례-대조군연구에서의 백신효과는 아래의 수식과 같이 표현된다.

$$\text{식 2: } VE=1-RR\approx 1-OR \text{ (단, OR≒RR인 경우)}$$

예를 들어서 다음과 같은 가상의 사례를 보자. 50만 명이 거주하는 한 지역사회에서 20만 명이 폐구균 예방접종을 하고 30만 명은 예방접종을 하지 않았으며, 1년 이후에 접종군에서는 100명, 비접종군에서는 600명의 폐구균 폐렴 환자가 발생했다고 하자.

표 5-1-1. 백신효과 측정을 위한 가상의 사례

	인구수	폐구균 폐렴 발생자 수	10만당 발생률
백신군	200,000	100	50
비백신군	300,000	600	200

위의 표에서 백신효과(*VE*)는 다음과 같이 계산된다.

$$VE=1-\frac{R_v}{R_u}=1-\frac{(100/200{,}000)}{(600/300{,}000)}=1-\frac{(50/100{,}000)}{(200/100{,}000)}=1-\frac{1}{4}=0.75$$

또는 위의 예에서 백신접종(노출)의 상대위험도는 (50/100,000)/(200,100,000)=1/4=0.25이므로 백신효과는 1-*RR*=1-0.25=0.75로 계산할 수 있다. 이를 %로 표현하면 백신효과는 75.0%가 된다.

3. 백신의 직접효과, 간접효과, 합계효과와 전체효과

지역사회에서 백신의 효과는 백신을 맞은 사람에게만 나타나는 것이 아니다. 백신을 맞은 사람에 대한 질병 예방효과를 직접효과(direct effect)라고 한다면, 같은 지역사회에서 백신을 맞은 사람들과 접촉하는 비백신군에게도 해당 질병에 대한 예방효과가 나타나는데 이를 백신의 간접효과(indirect effect)라고 한다. 이 둘을 합하여 합계효과(total effect)라고 하고, 인구의 일부 구성원이 백신을 접종함으로써 전체 인구가 경험하는 질병 또는 감염의 감소를 전체효과(overall effect) 또는 영향(impact)이라고 한다.

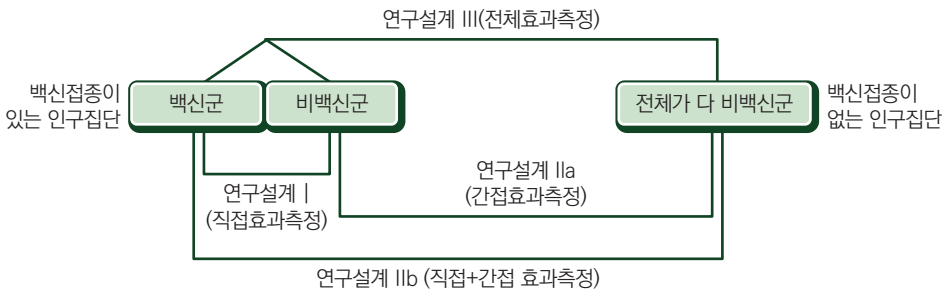


그림 5-1-1. 지역사회에서 백신의 직접효과, 간접효과, 합계효과, 전체효과를 측정하기 위한 연구설계

그림 5-1-1에서 보듯이 한 지역사회에서 백신군과 비백신군의 예방효과를 비교하는 것은 직접효과를 측정하기 위한 연구설계가 된다(연구설계 I). 백신의 간접효과를 보기 위해서는 백신접종사업을 수행한 지역사회의 비백신군과 백신접종사업이 아예 이루어지지 않은 지역사회의 인구(즉, 전체가 비접종군)의 발생률을 비교하여야 한다(연구설계 IIa). 이 둘의 효과를 합하면 합계효과를 얻을 수 있다(연구설계 IIb). 또한 백신의 전체효과를 측정하기 위해서는 백신접종사업이 수행된 인구집단과 그렇지 않은 인구집단 간의 발생률을 비교하여 얻을 수 있다(연구설계 III). 이 전체효과는 지역사회 예방접종사업의 실제 효과를 보여주는 지표가 되는데 전체효과를 측정하기 위해서는 각 인구집단의 백신 접종률이 필요하다. 이와 대조적으로 직접효과, 간접효과와 이들의 합으로 이루어진 합계효과는 개인수준의 접종여부를 알아야 측정가능하다. 이것이 합계효과와 전체효과의 차이이다. 여기서 알 수 있듯이 집단면역에 의한 백신의 간접효과, 합계효과, 전체효과를 측정하기 위해서는 해당 백신의 접종사업이 이루어지지 않은 대조 인구집단이 필요하다.

2 백신효과 측정을 위한 관찰역학 연구

백신효과 측정을 위한 관찰역학연구로 많이 사용되는 것은 코호트연구(cohort study), 사례-대조군연구(case-control study), 사례 검사-음성군 연구(case test-negative study), 사례-코호트연구(case-cohort study), 가구내접촉조사 연구(household contact study), 시계열분석연구(time-series analysis) 등이 대표적이다. 코호트연구와 사례-대조군연구 등 기본적인 역학연구설계의 종류와 특징에 대해서는 제1권 제2장 제2절의 역학연구설계방법을 참조한다.

1. 코호트연구

코호트연구는 백신군과 비백신군의 발생률을 비교하는 연구설계이며, 모든 백신효과 측정에서 가장 우선시되는 연구설계이다. 지역사회 백신효과를 측정하기 위한 코호트연구는 건강보험공단 자료 등의 데이터베이스를 이용하거나 또는 지역사회 유행자료를 활용하는 경우 등 후향적 코호트연구로 이루어지는 경우가 많다. 물론 전향적 코호트연구로도 측정이 가능하다. 연구방법은 우선 질병에 걸릴 수 있는 인구집단(population at risk)에서 백신 접종력에 따라 백신군과 비백신군을 선정하고 이들 각각의 발생률을 비교하는 것이다. 여기서 백신의 효과는 식 1을 사용한다(즉, VE=1-RR). 지역사회에서는 접종

일이나 관찰기간이 사람마다 다르기 때문에 인시(time-person)를 분모로 사용하여 발생률(즉, 발생밀도)을 계산하는 것이 바람직하다. 이때는 발생밀도로 상대위험도(*RR*)를 계산하여 백신효과를 측정한다. 만약 발생률을 누적발생률(cumulative incidence)이나 발병률(attack rate)로 측정을 했다면 상대위험도는 이를 이용해서 구하게 되고 이 상대위험도를 이용해서 백신효과를 구한다. 그러나 이는 개인 간 관찰기간이 다를 경우(즉, 개방 코호트인 경우) 발생밀도를 이용한 상대위험도보다 정확성이 떨어지게 된다. 실제 지역사회는 개인 간 관찰기간이 다 다르고 관찰단절(censoring)이 복잡하게 혼합된 인구집단이며, 각 개인의 연령이나 기저질환여부 등이 다르기 때문에 이들을 모두 고려한 콕스비례위험모형을 이용하여 위험비(hazard ratio, HR)로 상대위험도를 측정하는 경우가 많다. 당연히 이 경우도 백신효과는 *VE*=1-*HR*로 측정된다. 지역사회에서의 백신효과 측정은 이와 같은 이유로 콕스 모형을 많이 이용한다.

표 5-1-1은 누적발생률을 이용한 백신효과의 측정 사례라면 표 5-1-2는 발생밀도를 이용한 백신효과의 사례를 나타낸 것이다. 백신군과 비백신군에서 관찰된 총 인년으로 발생률을 계산하면 각각 100인년당 0.26명과 1.27명이고, 이때 RR=0.26/1.27=0.20이 된다. 따라서 백신 효과는 1-0.20=0.80 즉, 80%가 된다. 이때 백신효과의 95%신뢰구간도 같이 구하도록 한다.

표 5-1-2. 코호트연구에서 발생밀도를 이용하여 백신효과를 측정한 가상 사례

구분	총 인년	환자 수	10,000인년당 발생률
백신군	1,354,320	35	0.26
비백신군	658,646	84	1.27

표 5-1-3은 코로나 19가 유행 시 한 대학병원 근무자들의 백신접종 횟수에 따른 코로나 감염예방효과를 측정한 것이다. 근무자들의 관찰기관이 다 다르고 관찰중단이 적지 않으며, 백신효과에 영향을 주는 기저질환 등을 보정하여 콕스모형으로 보정위험비를 구한 결과이다. 이때 완전접종군의 예방효과는 1-[0.52(0.28-0.97)]=0.48(0.03-0.72)이 되어 48%(95% CI: 3-72)가 된다. 부스터까지 접종한 군은 같은 방식으로 54%(95% CI: 23-74)가 된다.

표 5-1-3. 코호트연구에서 위험비를 이용하여 백신효과를 측정한 사례

구분	보정 위험비(95% 신뢰구간)
비백신 또는 1차만 접종	1
완전 접종군(2차 접종)	0.52(0.28-0.97)

구분	보정 위험비(95% 신뢰구간)
부스터 접종군(3차 접종)	0.45(0.26-0.77)

최근에 예방접종 데이터베이스와 건강보험 데이터베이스를 연계해서 코호트연구설계로 백신효과를 측정하는 연구가 활발하게 이루어지고 있는 것은 바람직하지만, 2차 자료를 이용한 연구에서는 자료의 특성에 대한 깊은 이해와 엄밀한 연구설계, 그리고 이에 맞는 분석방법이 뒷받침되어야 한다. 예를 들어서 소아의 폐구균예방접종의 효과를 평가하기 위하여 우리나라 예방접종 DB와 건강보험자료를 연계하여 분석했을 때 오히려 접종군에서 폐렴이나 중이염 등으로 인한 의료이용이 더 많이 관찰되는 역설이 보고되기도 했는데 이는 비접종군의 의료이용 행태가 접종군과 크게 다르기 때문이다(관련문헌: Sohn et al., 2021).

2. 사례-대조군연구

환자-대조군연구라고도 하는 사례-대조군연구(case-control study)는 질병군과 비질병군을 먼저 구분한 후 이들이 과거에 백신을 맞았는지를 조사하는 역방향(backward directionality) 연구이다. 코호트연구가 백신여부를 먼저 관찰하여 백신군과 비백신군을 나누고 시간에 따른 발생률을 구하는 정방향(foward directionality) 연구인 것과 반대 특성을 갖는다. 이러한 특성 때문에 사례-대조군연구에서는 코호트연구와 같은 발생률 측정이 불가능하고, 오직 사례군과 대조군에서의 과거 접종률만 알 수 있게 된다. 반면 사례-대조군연구도 감시 프로그램에 전향적으로 통합되어 시간이 지남에 따라 사례군와 대조군을 등록하는 전향적 연구(prospective study)로 진행할 수 있고 실제로 많은 전향적 사례-대조군연구가 수행되고 있다. 사례-대조군연구는 모두 후향적 연구(retrospective study)라는 주장은 역학연구설계에서 방향성(directionality)과 시간성(timing)을 구분하지 못하는 데서 오는 오류이다. 사례-대조군연구는 시간성 기준 전향적 연구와 후향적 연구가 모두 가능하다. 다만 방향성에서는 반드시 역방향만 가능하다(제1권 제2장 제2절의 역학연구설계방법 참조).

사례-대조군연구의 타당성은 사례정의(case definition)과 대조군의 선택(control selection)에 크게 좌우된다. 백신효과 측정을 위한 사례-대조군연구에서 좋은 대조군이 가져야 할 특성은 다음과 같다.

- 사례군과 대상 병원체에 노출될 위험이 같아야 한다.
- 백신 접종 전에 해당 질병에 유사하게 취약해야 한다.
- 백신 접종 상태와 관계없이 대조군이 모집되어야 한다.

- 의료이용 및 백신 접종에 대해서 사례군과 동일한 접근성을 가져야 한다.

사례-대조군연구에서의 백신의 효과는 식 2를 이용한다(즉, $VE=1-OR$). 표 5-1-4는 폐구균 백신 효과를 측정하기 위하여 여러 기관이 참여한 병원기반 사례-대조군연구 결과의 한 부분이다. 여기서 65~74세 군에서는 백신의 효과가 57.4%(95% CI: 19.4-77.5)로 유의하나, 75세 이상에서는 백신효과가 6.7%(95% CI: -73.9-49.9)로 통계적으로 유의하지 않은 것을 알 수 있다.

표 5-1-4. 폐구균 백신의 침습성 폐구균질환의 연령별 예방효과 측정 사례

구분	보정 오즈비(95% 신뢰구간)
65~74세	0.436(0.225-0.706)
75세 이상	0.933(0.501-1.739)

3. 사례 검사-음성군 연구

사례 검사-음성군 연구(case test-negative study)는 지역사회 백신효과 측정에서 가장 흔하게 사용되는 방법 중의 하나이다. 이 연구설계가 가장 흔하게 이용되는 인플루엔자 백신효과 측정을 예로 들어서 그림을 그리면 그림 5-1-2와 같다.

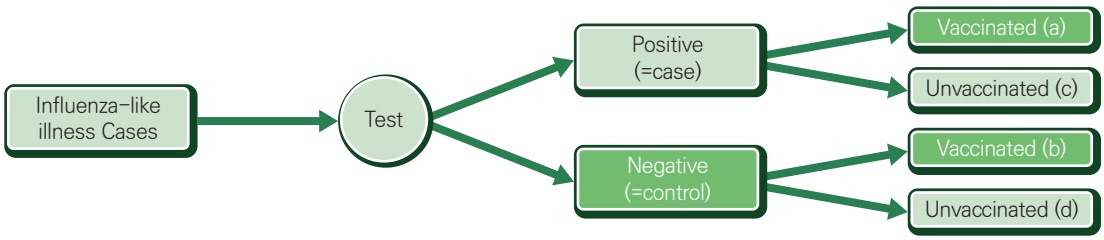


그림 5-1-2. 사례 검사-음성군 연구를 이용한 인플루엔자 백신의 효과 측정연구의 설계

이 연구설계는 그림 5-1-2에서 보는 것처럼 의료기관을 방문한 인플루엔자 유사환자(influenza-like illness, ILI)를 대상으로 인플루엔자 검사(신속항원검사 혹은 PCR)를 수행해서 양성군과 음성군으로 나누고, 다시 양성군과 음성군의 백신여부를 조사하여 오즈비를 구하여 백신효과를 측정하는 것이다. 인플루엔자 유사환자는 발열과 함께 기침 또는 인두통(sore throat)을 가진 사람으로 정의되기 때문에 인플루엔자 이외에 다양한 호흡기 바이러스들이 같은 증상을 일으킬 수 있다. 따라서 이 방법의

핵심가정은 인플루엔자 백신이 유사한 증상을 가진 다른 호흡기 질환에 대해서 교차보호 효과가 없다는 것이다. 이 연구설계에서는 발생률을 구할 수 없고 OR만 구할 수 있으므로 1-OR이 백신효과로 측정된다.

사례 검사-음성군 연구는 사례군과 유사한 증상을 가진 사람 중에서 대조군이 선정되므로 일반적인 사례-대조군연구에서처럼 두 군간 치료접근성과 의료이용행태 차이로 인한 교란효과가 적어지는 장점이 있다. 또한 사례 검사-음성군 연구는 감시체계를 수행하는 기관에서 같이 진행할 수 있기 때문에 미국 등 여러 국가에서는 이 방법을 이용한 인플루엔자 예방접종효과를 모니터링하고 있다.

표 5-1-5. 병원기반 인플루엔자 감시체계에서 인플루엔자 예방접종 효과 측정을 위한 사례 검사-음성군 연구

구분	양성군	음성군	오즈비
백신접종 있음	263	275	0.552
백신접종 없음	90	52	1

표 5-1-5는 2016/17 절기에 우리나라 병원기반 인플루엔자 감시체계에서 사례 검사-음성군 연구를 수행한 결과의 일부이다. 여기에 참여한 각 병원에 내원한 ILI환자에 대해서 신속항원검사를 수행하여 양성군과 음성군으로 나눈 후 각각에 대해서 백신접종여부를 확인한 결과이다. 이 연구에서 OR=0.552이므로 이때 인플루엔자 백신의 효과는 1-0.552=0.448 즉, 44.8%가 된다. 인플루엔자 백신의 경우는 매년 백신과 유행바이러스가 달라지므로 백신효과의 모니터링은 필수이며 미국의 많은 병원으로 구성된 다기관에서 이 방법으로 인플루엔자 백신효과를 모니터링하고 있다.

4. 가구접촉조사 연구

가구접촉조사 연구(household contact study)는 전파확률측정 연구(transmission probability study)라고도 한다. 이 방법은 사례가 발생한 가구를 하나의 미니 코호트로 간주하고 각 가구 내에서 2차발병률(secondary attack rate)을 측정하여 비교하는 방법이다. 즉, 가구 내 1차 발병자와 접촉한 가구원 중 백신을 맞은 가구원에서의 2차발병률과 백신을 맞지 않은 가구원에서의 2차발병률을 분석하는 방법이다. 이를 공식으로 표현하면 식 3과 같다.

$$식\ 3:\ VE=\frac{SAR_v}{SAR_u}=1-\frac{\frac{\text{vaccinated infections}}{\text{vaccinated contacts}}}{\frac{\text{unvaccinated infections}}{\text{unvaccinated contacts}}}$$

식 3에서 VE는 백신효과, SAR_u와 SAR_v는 각각 비접종군과 접종군에서의 이차발병률을 의미한다. 이 연구를 위해서는 가구 내 접촉과 전파에 대한 엄밀한 역학조사가 있어야 한다. 표 5-1-6은 5개의 가구에 대해서 조사한 가상의 사례로 1차 사례와 접촉한 가구원을 접종여부에 따라 나눈 후 발병여부를 관찰한 것이다. 여기서 백신접종 가구원은 모두 8명인데 이 중 1명이 발생하여 이차발병률은 1/8=0.125가 되고, 비접종군에서는 8/12=0.667이 된다. 이때 공식을 이용하면 VE=1-(0.125/0.667)=0.813, 즉 81.3%가 된다. 이 방법은 백신의 전파예방효과를 측정할 수 있으며 위스콘신과 테네시 주 등 지역사회에서 수행한 연구들이 최근에도 계속 보고되고 있다.

표 5-1-6. 가구접촉조사 연구의 가상사례

가구번호	발병자 수/백신접종가구원 수	발병자 수/비접종가구원 수
1	0/2	2/2
2	1/3	2/3
3	0/0	1/2
4	0/2	1/2
5	0/1	2/3
합계	1/8	8/12

5. 사례-코호트연구

사례-코호트연구(case-cohort study)는 감시체계 자료를 활용하는 방법으로 스크리닝법(screening method)이라고 부르기도 한다. 이 방법은 인구전체(즉, 코호트)의 예방접종률을 알고 있을 때 이 코호트 중 일부에서 발생한 환자(즉 사례군)에서의 예방접종률과 비교하는 방법이다. 이를 공식으로 표현하면 식 4와 같다.

$$식\ 4:\ VE=1-\left(\frac{PCV}{1-PCV}\right)/\left(\frac{PPV}{1-PPV}\right)$$

식 4에서 *VE*는 백신효과, *PCV*와 *PPV*는 각각 사례군과 전체 인구집단에서의 접종률이다. 사례-코호트연구에서 계산되는 것도 OR이기 때문에 식 4가 성립이 된다. 이 연구에서 고려해야 하는 점은 인구집단의 연령별로 백신접종률이 다르므로 연령별로 층화하여 비교하여야 하며 이와 같이 다른 제3의 요인들을 보정하기 위해서는 이 요인으로 층화하여 분석해야 하기 때문에 교란변수를 보정하는 것이 쉽지 않은 것이 단점이다. 또한 인구집단의 접종률이 매우 높거나 낮을 때 이 방법으로 측정한 백신효과는 참값에서 멀어질 수 있다는 것도 단점이 된다. 그러나 인플루엔자 백신처럼 인구집단별 접종률이 잘 알려져 있을 때 인플루엔자 환자중 접종률만 조사하면 바로 계산이 가능하기 때문에 거의 실시간으로 백신효과를 모니터링할 수 있다는 장점이 있어서 유럽연합에서는 I-MOVE 프로젝트를 통해서 이 방법으로 인플루엔자 백신효과를 모니터링하고 있으며 그 결과는 매년 Eurosuveillance에 발표되고 있다.

6. 시계열분석 연구

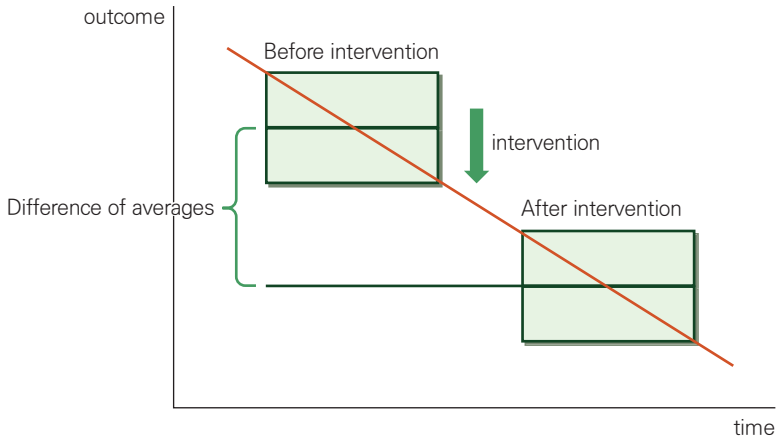


그림 5-1-3. 특정 중재사업(예를 들면 예방접종사업)의 효과평가에서 단절적 시계열분석이 필요한 이유

시계열분석 연구(time-series analysis)는 일종의 생태학적 모델링 연구이며 다양한 시계열분석 모델링 기법이 사용된다. 여기서는 그중의 하나인 단절적 시계열분석법(interrupted time-series analysis)만 소개한다. 단절적 시계열분석은 특정시점에 수행된 어떤 중재사업의 효과를 측정하기 위해 개발된 방법으로 예방접종의 경우, 특정 예방접종사업이 시작되었다면 이 이전과 이후에 해당 질병의 발생정도를 비교하여 백신효과를 평가하는 방법이다. 단순하게 백신접종사업 전후의 발생률을 비교하는 것은 그림 5-1-3과 같은 오류가 따른다. 그림에서 중재 전후의 평균만 비교하면 차이가 있는 것처럼 보

이지만 실제로는 중재 이전부터 계속 발생이 감소하고 있었기 때문에 실제로는 중재가 없더라도 이 평균의 차이는 관찰되었을 것이다. 따라서 단순한 예방접종사업 전후의 평균 발생률 비교만으로는 예방접종사업의 효과를 측정할 수 없고, 이때 사용되는 것이 단절적 시계열 분석이다.

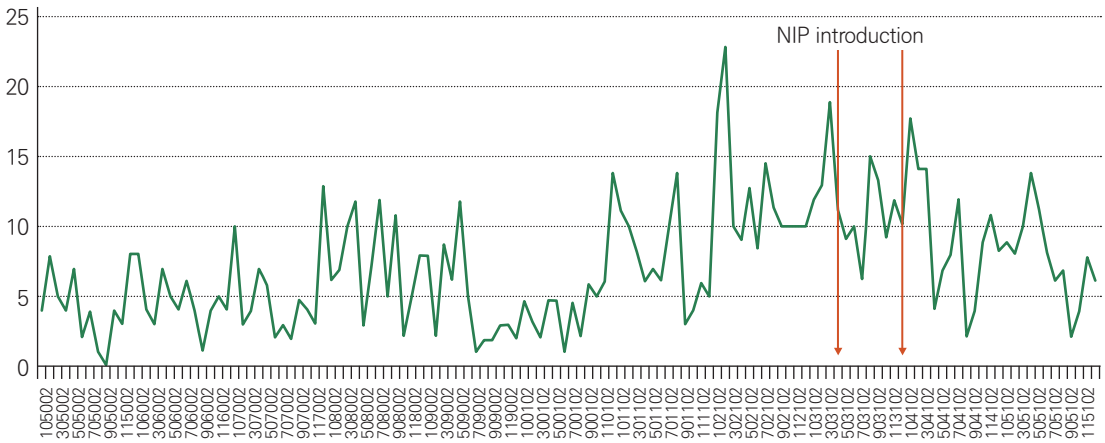


그림 5-1-4. 65세 이상 인구에서 폐구균으로 인한 패혈증의 월별 발생률(2005~2015년)

표 5-1-7. 2013년 65세 이상 인구에 PPV 접종 도입 후 폐구균패혈증의 발생률 변화에 대한 단절적 시계열 분석 결과

	Coefficient	Std. Error	t value	Pr (>|t|)
Intercept β_0	3.342	1.032	3.24	0.0016
Baseline trend β_1	0.061	0.018	3.29	0.0013
Level change β_2	2.915	2.215	1.32	0.1908
Trend change β_3	−0.3342	0.1424	−2.35	0.0206

출처: 천병철. 주요 폐렴구균 감염증 발생추이 및 폐렴구균단백결합 백신의 유용성 평가. 질병관리본부 보고서. 2016

단절적 시계열 분석의 구체적인 사례를 들면 다음과 같다. 그림 5-1-4는 2005년 1월부터 2015년 12월까지 월별로 폐구균 패혈증의 발생상황을 건강보험자료에서 추출한 것이다. 그런데 2013년부터 65세 이상 인구에 폐구균접종(PPV)이 국가예방접종사업에 포함되었다. 그러면 이 접종사업이후에 폐구균으로 인한 패혈증이 감소하였을까? 이 자료는 시간적 자기상관(계절성)이 있는 자료이므로 ARIMA를 이용한 단절적 시계열분석을 하였을 때 표 5-1-7과 같은 결과를 얻을 수 있다. 여기서 β1은 접종사업 도입 이전의 발생경향을 나타내는 것으로 계수가 양수로 유의한 것을 알 수 있다. 이는 백신사업도입 이전에 노인에서 폐구균패혈증이 증가하고 있었음을 의미한다. 같은 표에서 β3는 접종사업

도입 이후의 발생경향으로 음수로 유의한데 이것은 백신도입 이후 폐구균 패혈증이 감소추세로 변한 것을 의미한다. 즉, 백신사업의 도입은 노인에서 폐구균패혈증의 발생추이를 감수추세로 변화시킨 것이다.

3 맺음말

여기서는 백신의 직접효과, 간접효과, 합계효과, 전체효과에 대한 정의와 이를 측정하기 위한 연구설계를 기술하였고, 구체적인 백신측정 연구설계로 코호트연구, 사례-대조군연구, 사례 검사-음성군 연구, 사례-코호트연구, 가구접촉조사 연구, 시계열분석을 이용한 모델링 등을 사례와 함께 소개하였다. 이 밖에도 간접코호트연구(indirect cohort study, Broome method) 등의 다양한 관찰역학 연구 방법이 백신효과 평가에 사용된다.

백신효과의 측정은 공중보건학적으로 백신 효능의 측정보다 중요하다. 그러나 백신 효능은 백신판매 허가를 위해서 반드시 필요하므로 광범위하게 수행되나 실제 판매 이후에 지역사회효과의 측정은 선진국을 제외하고는 잘 수행되지 않고 있다. 미국이나 유럽연합에서는 인플루엔자처럼 매년 백신과 유행 바이러스가 바뀌는 경우 다양한 기관에서 감시와 함께 백신효과를 모니터링하고 있으며, 다양한 백신효과 평가사업을 하고 있다. 한국의 경우는 예방접종등록자료와 건강보험자료를 연계하여 2차 자료를 활용한 백신효과평가가 수행될 수 있다는 장점이 있으며 실제로 여러 편의 논문이 출판되었다. 그러나 이는 연속적이지 않은 용역이나 연구의 결과물이며 체계적인 백신효과 평가가 아직 잘 이루어지고 있지는 않다. 매년 막대한 예산으로 수행하는 예방접종사업의 효과를 평가하고 모니터링하는 것은 감염병관리와 우리나라의 백신사업 선진화를 위해서 시급한 과제 중의 하나이다.

참고 문헌

1. 천병철. 관찰연구를 통한 백신효과의 평가. 2025년 고려의대 백신혁신센터 전문가 교육과정 자료집. 2025.
2. 천병철, 최승아, 유석현 외 역. 중급 역학. 4판. 서울: 범문에듀케이션. 2024.
3. ECDC. Lecture series. Applied Epidemiological Research for Vaccine Preventable Diseases: Vaccine Effectiveness. 2012.
4. Farrington CP. Estimation of vaccine effectiveness using the screening method. Int J Epidemiol. 1993;22(4):742-6.
5. Halloran ME, Longini IM, Struchiner CJ. Design and analysis of vaccine studies. 2010th ed. New York, NY: Spring-

er; 2009.

6. Halloran MEH, Longini IML, Struchiner CJ. Design and Interpretation of Vaccine Field Studies. Epidemiological Reviews. 1999;21:73-87.
7. Rodrigues LC, Smith PG. Use of the case-control approach in vaccine evaluation: efficacy and adverse effects. Epidemiol Rev. 1999;21(1):56-72.
8. Sohn S, Hong K, Hwang H, et al. Paradoxical health care utilization patterns among children in Korea who did not receive mandatory pneumococcal vaccination. Vaccine. 2021;39(7):1096-1100.

제2절

예방접종 의향에 대한 이해와 소통 중심의 대응 방안

1 배경

1. 공중보건의 목표 실현과 효과적인 건강 위험 소통

공중보건은 인구집단의 건강을 증진(promotion)하고, 질병과 상해를 예방(prevention)하며, 건강 위험으로부터 취약층과 지역사회를 보호(protection)하는 것을 핵심 목표로 삼는다. 이런 목표를 실현하기 위해서는 여러 요건이 필요하다. 과학적 근거에 기반한 정책 결정이 가능하도록 관련 연구개발(R&D)이 필요하고, 건강에 긍정적인 환경을 조성하는 제도와 정책 개발과 주요 이해관계자 사이의 협력적인 거버넌스 등도 필요하다. 그러나 건강할 권리의 주체이자 공중보건 정책과 서비스의 수용자면서 이용자인 국민, 곧 공중(the public)의 참여와 협력 없이는 이러한 목표 달성은 어렵다. 공중보건이 인간의 심리, 의사결정, 그리고 행태를 중요하게 다뤄온 이유이기도 하다.

최근에는 개인과 집단의 건강 의향과 결정 및 행태에 소통(커뮤니케이션)이 미치는 영향력이 크게 높아졌다. 이는 정보통신 기술과 매체의 발전 결과로, 스마트폰 등의 디바이스를 통해 뉴스와 정보의 빠른 확산, 소셜 미디어를 통한 정보 접근성 향상, 특정 선호와 가치를 공유하는 온라인 커뮤니티의 영향력 등 과거와는 완전히 다른 소통 환경이 펼쳐지게 되었다. 이는 지역사회와 인구 집단의 건강한 행태를 취하도록 목표하는 공중보건이 소통의 환경과 특성을 이해하지 않고서는 제대로 효과를 낼 수 없게 되었음을 의미한다.

효과적인 소통은 사실상 공중보건의 핵심 기능이자 전략이다. 공중보건이 개인이 아닌 지역사회 차원의 위험(community risks)을 다루고, 잠재적 위험(potential threats)에 대해 다수의 공중(large population)에게 주의를 당부하고 예방을 강조해야 한다는 점에서, 소통의 중요성은 더 크다 할 수 있다. 미국의 『Healthy People 2020』과 국내 『신종감염병 대유행 대비 중장기 계획』 등은 효과적인 소통을 세부 목표이자 핵심 과제로 포함하고 있는데, 공중보건의 목표 달성을 위해 소통의 강화가 주요 전략임을 엿보게 한다.

2. 소통 중심의 공중보건 접근법

리말(Rajiv Rimal) 등은 "왜 공중보건에서 소통이 중요한가?"라는 글에서 건강 행태에 영향을 미치는 환경, 사회, 심리적 요인이 과거보다 더 복잡하고 불확실해졌다고 지적하면서, 기존의 접근법에 소통 과학의 이론과 방법론을 적용함으로써 더 효과적으로 건강행동을 이해하고 중재를 이끌 수 있다고 주장한다.

여기서 소통은 기본적으로는 그림 5-2-1처럼 두 주체 간의 언어적·비언어적 정보 교환을 의미하지만 여기에 한정되지 않는다. 주체의 상호작용을 통해 공유되는 '의미'와 '상징' 역시 소통 결과에 유의미한 영향을 미치기 때문이다. 따라서 소통 중심의 공중보건 접근법은 누가(전달자와 수신자), 무엇을(메시지), 어떻게(채널) 교환하는가의 구조만이 아니라, 보건당국, 전문가, 산업, 제공자와 환자 단체 등 다양한 주체들이 일반 공중과 함께 맺는 '관계의 맥락(relational contexts)' 속에서 건강에 대해 어떤 의미를 공유하고, 그로부터 어떤 결과가 이끌어지는지의 '과정'을 함께 중요하게 다룬다.

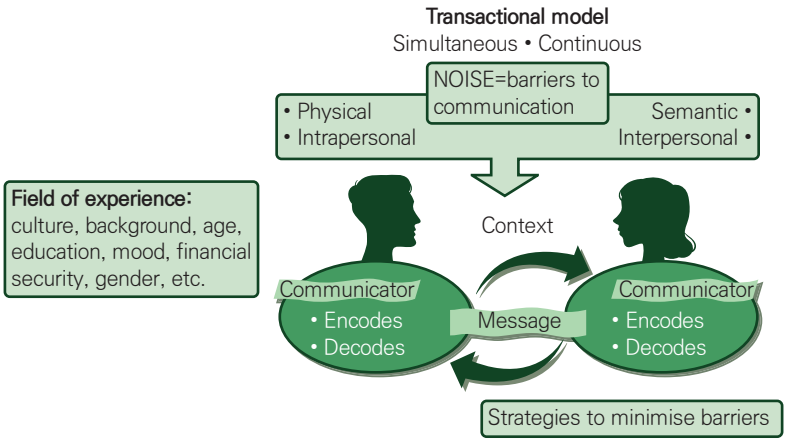


그림 5-2-1. 소통의 정보 교환 모델(Transactional Model of Communication)

또한, 소통 중심의 접근법은 언제나 소통을 방해하는 다양한 물리적, 비물리적 장벽(barriers to communication)이 있고, 소통이 벌어지는 상황적 맥락(context)이 중요하다고 강조한다. 예방접종을 예로 들면, 어느 지역사회에 예방접종을 둘러싼 물리적, 사회심리적 장벽이 무엇인지, 지난 시기 또는 다른 백신 예방접종과 관련하여 벌어진 긍·부정적인 사안이 무엇인지를 파악하고 현재의 소통 장벽을 줄이는 전략을 예방접종 사업의 성공을 위한 필수 요건으로 강조하게 된다.

끝으로, 공중보건에 적용할 수 있는 소통 접근법은 하나 이상임을 이해할 필요가 있다. 널리 알려

진 건강 소통(Health Communication) 외에도 위험 소통(Risk Communication), 위기 소통(Crisis Communication), 과학 소통(Science Communication), 정책 소통(Policy or Political Communication) 등 분야별로 축적된 연구와 실무 전략을 공중보건에 활용할 수 있다. 또는, 소통의 각 분과 학문의 개념을 활용할 수도 있다. 저널리즘 연구가 다루는 효과적인 메시지 프레이밍(message framing), 광고홍보나 소셜 마케팅 연구가 강조하는 '맞춤'의 전략들(청중-매체-메시지 맞춤 등), 아니면 정치 커뮤니케이션이 강조하는 위기 소통의 책무성 등은 결코 공중보건과 거리가 먼 주제가 아니기 때문이다.

3. 코로나19 대응의 교훈: "위기대응의 모든 것은 사실상 소통이다."

코로나19 팬데믹은 재난적 공중보건 상황 대응에서 효과적인 소통이 얼마나 중요한지를 입증한 중요한 사례이다. 감염의 대확산을 차단하려면 감염병 유행에 맞설 행정적 대응 체계를 구축하는 것만큼, 전 국민이 신속히 예방 행동을 취해야 하는데, 이를 위해서는 위험 정보의 신속, 정확한 제공, 행동 지침의 효과적인 전달 등 소통의 기능과 체계가 원활하게 작동해야만 한다. 최근 Lazarus 등은 전문가 델파이 조사를 통해 미래 감염병 위기 대비·대응을 위한 6가지 핵심 권고를 제시했는데, 이때 '소통 향상'은 '지역사회 참여 강화', '신뢰 확보'와 함께 전체 참여자의 절대적 수준의 동의를 이끈 3대 항목의 하나였다.

문헌들은 코로나19 같은 공중보건 위기대응에 효과적인 소통이 중요한 이유를 다음과 같이 설명한다. 첫째, 효과적인 소통은 대규모의 예방 행동을 조직화하는 것에 기여한다. Bavel 등은 감염병 대유행 시 생명과 건강 보호를 위해 전격적인 행동 변화가 일어나야 하는데, 위기는 그 자체로 개인에게 상당한 심리적 부담을 의미하고 의사결정의 불확실성을 높이기 때문에 감염 정보가 제공된다고 해서 즉각적 행동 변화를 가정하기 어렵다고 지적한다. 그러면서, 이때 가장 효과적인 방식으로 정보와 권고를 전달하고 실천을 돕는 소통은 절대적으로 중요하다고 설명한다. 둘째, 효과적인 소통을 통해 부정확하거나 왜곡된 정보의 영향, 즉 인포데믹을 예방할 수 있다. 감염병 위기가 불러온 불안과 공포는 정보 탐색과 의존도를 높이는데 동시에 팬데믹의 높은 뉴스 가치로 정보의 범람 상황이 초래될 수 있다. 바로 이 과정에서 불분명한 출처, 부정확한 내용, 또는 의도적으로 왜곡된 정보에 공중의 노출 가능성이 커진다. 따라서 정확성, 투명성, 개방성 등을 원칙으로 하는 소통을 통해 이런 문제의 위험으로부터 취약층을 보호하는 기능을 발휘할 수 있다.

셋째, 팬데믹 상황에서 미디어를 통해 전달되는 메시지는 그 의도와 상관없이 노출된 사람의 정신건강에 부정적인 영향을 미칠 수 있기 때문에 효과적인 소통이 중요하다. 감염의 출처, 진원지, 감염된 환

자에 대한 수많은 정보 속에서 무분별하게 만들어지고 유통되는 서사(narratives)는 누군가에게 심각한 낙인이나 차별을 유발할 수 있다. 그동안 문헌은 코로나19팬데믹에서 감염 환자가 죄책감과 두려움을 느끼고, 사회활동이 제한된 청년층이 무력함과 고립 속에 어려움을 겪고, 가족을 만나기 어려운 요양 시설 노인과 그 가족들이 슬픔과 괴로움을 느낀다고 보고했다. 특히나, 감염병 대응 현장의 공중보건 인력들 사이 높은 스트레스와 위험 인식이 알려졌다. 이때 옹호(advocacy)와 지지(support)를 강조하고 공감과 연대를 이끄는 소통은 사회의 회복에 유의미한 역할을 할 수 있다. 끝으로, 소통은 기본적으로 양방향이며, 전달만큼 '청취'를 중심에 둔다. 그런 점에서, 긴급성을 이유로 위기대응이 일방적이거나 권위적으로 되지 않도록 소통의 역할을 발휘할 수 있다. 특히, 개인과 집단의 참여와 협력을 강조하고, 건강 정보의 접근과 활용에 가로놓인 격차에 주목하면서 정보 취약층을 옹호하고 지원하는 활동을 벌일 수 있다.

2 예방접종 의향과 행태 이해

1. 백신과 예방접종

예방접종은 역사적으로 가장 성공적인 공중보건 사업이자 성과로 지적된다. 예방접종을 통해 획득한 면역력은 아동을 비롯한 건강 취약층을 감염성 질병으로부터 보호하고, 지역사회나 국제사회가 감염병 유행에 직면할 때 이를 차단할 가장 강력한 수단은 지역 주민 다수의 예방접종을 통한 집단 수준의 면역력 획득이라고 알려진다.

이때의 예방접종(vaccination)은 백신을 제공하는 측, 즉 관련 정책이나 사업의 이름이기도 하지만 백신에 대한 개인들의 의향과 의사결정, 그리고 귀결되는 행태로 보는 것이 중요하다. 이런 개인의 의향과 의사결정(vaccine related intentions and decisions)은 백신의 수용(vaccine acceptance)으로 나타나고 행태적 결과인 예방접종 참여(vaccine uptake, vaccination)로 이어진다. 인구 집단의 접종 참여는 예방 접종률(uptake rates)의 이름으로 국가 백신 정책과 사업의 핵심 지표로 고려된다.

그러나 이와 같은 백신 접종은 코로나19 팬데믹을 거치며 도전을 맞이하게 된다. '과학적 혁신'으로 불렸던 빠른 코로나19 백신의 개발은 긴급 사용승인 등을 거쳐 우리나라를 포함, 다수 국가에 보급되었으나, 물리적으로는 예방접종이 가능한데도 접종을 기피하거나 주저하고, 나아가 반대하는 의향 및 행태가 국제적인 문제로 등장한 것이다. 실제 세계보건기구는 코로나19 예방접종에 대한 주저나 거부

를 중대한 공중보건 위협으로 발표하기도 했다.

이러한 격차, 즉 백신의 안전성과 효과성을 알리면서 접종을 권고하는 보건당국, 이동과 모임을 제한하는 사회적 거리두기 등의 조치에 비해 예방접종은 더 비용-효과적일 수 있다고 설명하는 전문가 집단이 실제 공중의 수용 및 행태와 동일하게 나타나지 않는 인식과 행태의 격차는 공중보건의 성찰과 더 정교한 중재를 요청한다.

이 과정에서 백신 소통의 역량과 효과성 강화가 중요 대안으로 지목된다. 개인의 예방접종 의향과 행태는 '진공 상태'에서 형성되지 않기 때문이다. 실제로 문헌들은 코로나19 백신 수용에 성별, 연령, 기저질환 여부 등 기존에 알려진 영향 요인 외 코로나19 백신의 안전성과 효과성, 접종 이후 부작용과 이상반응, 및 보건의료 체계에 대한 불신 등이 공통적인 우려가 깊숙이 연관되어 있다고 지적한다.

2. 코로나19 예방접종을 통해 드러난 우려와 문제점

국내 코로나19 예방접종은 2021년 2월 의료 종사자, 면역 저하자 등 고위험 층을 우선 대상으로 시작하여 8월부터 전 국민 대상으로 이루어졌다. 초기 기초 접종에서는 1차 접종 87.9%, 2차 접종 87.1%의 높은 참여율을 기록했고, 이후 3차 접종 65.6%, 4차 접종 14.7%로 보고된다. 2024년 현재는 '23~24절기 접종', '24~25절기 접종' 등 계절 감염 예방 접종 사업의 일환으로 시행되고 있으며, 65세 이상, 12~64세 면역저하자, 감염취약시설 구성원 등 고위험군을 적극 권고 대상으로 하며, 이들 고위험군에게는 인플루엔자와 코로나19 백신의 동시 접종이 권고되고 있다. 본 절의 초점인 수요자 또는 당사자 중심의 관점에서 우리 사회의 백신 소통 역량과 효과성 강화를 위해 함께 짚어볼 내용은 다음과 같다.

첫째, 달라진 소통의 맥락을 고려한 코로나19 예방접종 소통 전략이 필요하다. 지난 시기 코로나19 예방접종은 팬데믹의 위기를 극복할 국가적 대안으로 강조되었고, 개인의 동기는 물론 공동체를 위하는 이타적 보호 동기가 크게 작용했다고 볼 수 있다. 그러나 현재는 그와 다른 맥락이다. 즉 현재는 일상에서 자기 자신의 건강을 관리하는 방법의 하나로 감염 취약층이 코로나19 백신을 고려하고, 수용과 접종 참여가 이루어지도록 돕는 것이 필요하다. 따라서 '그때는 모두의 위기였고 모두의 접종이 필요한 상황이었지만 지금은 코로나19가 그렇게 심각한 상황이 아닌 것 같은데 접종이 꼭 필요한가?' 또는 '아직도 안전한지 확신이 안 서는 반면 이상 반응은 여전히 우려되는데 반드시 접종해야 하는가?' 최근에는 '동시 접종을 권고하는데 오히려 더 위험한 것은 아닌가? 등 예상되는 의문에 가장 효과적인 내용과 방식으로의 소통을 위한 새로운 맞춤 전략과 계획이 필요하다.

둘째, 백신 신뢰를 높일 방안이 필요하다. 이때의 백신 신뢰는 약물이나 과학 신뢰만이 아니며, 백신

을 도입하고 예방접종 사업을 추진하는 공적 기관 등 제도적 신뢰가 중요하다. 그동안 국내에서는 예방접종 관련 정부 정책[예: 접종증명·음성확인제도(방역패스), 코로나19 예방접종 피해와 관련한 국가보상제도]을 둘러싼 논쟁과 갈등이 적지 않았다. 코로나19 백신에 대해서는 공적 주체를 향한 신뢰가 예방접종의 의향과 행태 결정에 영향을 미친다는 문헌의 보고가 이어지고 있다는 점에서, 백신 신뢰의 향상 특히 공적 신뢰를 높일 방안이 필요하다.

셋째, 코로나19 예방접종이 연례화되고, 65세 이상 노인과 기저질환자 등 취약층에게 코로나19 백신과 인플루엔자 백신의 동시 접종이 권고되는 가운데, 국내 코로나19 예방접종에 관련된 경험과 태도의 현황을 볼 때, 세계보건기구 등의 권고가 실제 공중의 수용에서는 동일하게 나타나지 않을 수 있음을 가정하고, 동시 접종 수용 의향 및 실제 참여를 높일 효과적인 지원과 촉진의 방안을 마련할 필요가 있다.

이러한 내용들은 '제공' 중심의 접근이 아닌 '수요' 중심, '당사자' 중심의 접근을 요구한다고 볼 수 있다. 특히, 백신 수용에 관련된 개인의 의향과 경험을 심도 있게 이해할 필요가 있다.

3. 백신 주저

예방접종의 의향과 의사결정에서 중요한 개념으로 '백신 주저(Vaccine Hesitancy)'가 있다. 가장 널리 알려진 백신 주저의 정의는 '예방접종이 가능한데도 접종을 미루거나 기피'하는 것이며, 최근에는 백신 주저를 동기의 영역(motivation domain)으로 보고, 개인이 예방접종에 대해 갈등을 느끼거나 반대하려는 동기가 있는 상태로 이해할 것을 강조한다. 이때 백신 주저는 예방접종 의향(intention)과 의지(willingness)를 모두 포괄하며, 실제 행동 결과인 예방접종 참여(vaccination, uptake)와 독립된 변수로 행태를 예측하는 요인으로 다룰 수 있다.

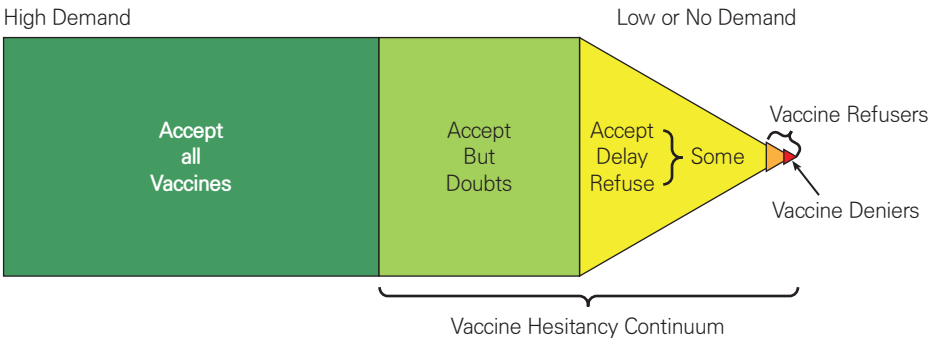


그림 5-2-2. 백신 주저 연속체(Vaccine Hesitancy Continuum)

문헌들은 백신 주저를 복잡한 현상이며, 시간이나 장소, 또한 백신 종류별로 다르게 나타날 수 있고, '연속체'의 개념으로 볼 것을 강조한다(그림 5-2-2).

이 연속체는 모든 백신을 수용하는 입장(Accept All Vaccines)부터 모든 백신을 거부하거나 반대하는 사람(Vaccine Refusers, Vaccine Deniers)까지 포함하고 그중 백신 주저는 '백신을 수용하나 의구심을 표하거나(Accept but Doubts)', '접종 수용(accept), 지연(delay), 거절 (refuse)'이 '어느 정도(some)' 있는 상태에 해당한다. 즉, 백신 주저는 단순히 백신 수용 여부의 이분법 이상의 다수 유형으로 존재할 수 있다.

4. 설문조사 결과로 본 코로나19 백신 수용 의향

현재 백신 주저를 다루는 '표준 질문'은 없다고 볼 수 있다. 문항에 따라 백신 주저의 현황이 서로 다를 수 있고 따라서 단순 비교가 어렵다. 이를 유념하면서 국내에서 시행된 설문 조사 연구가 제시한 현황을 보면 다음과 같다.

1) 접종 사업 전

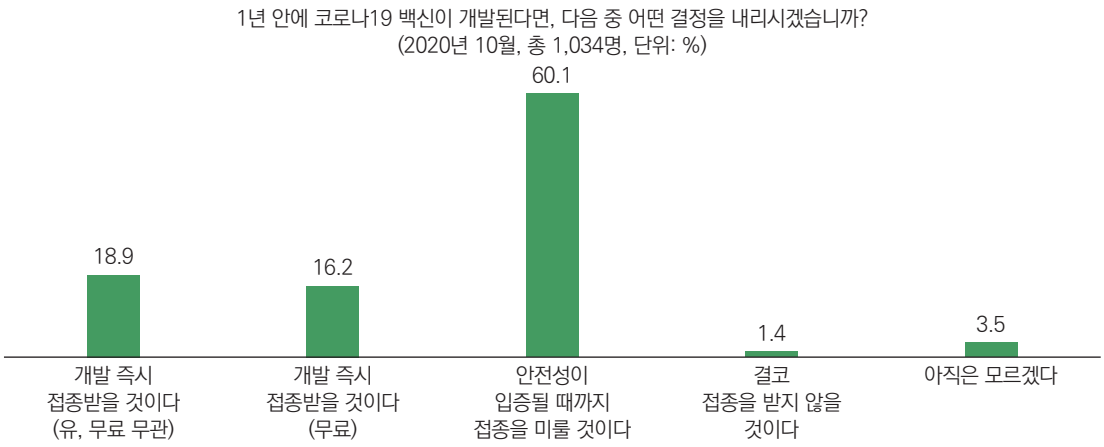


그림 5-2-3. 코로나19 백신 수용 의향(2020년)

국내 코로나19 백신이 도입되기 전(2020년 10월 만 18세 이상 국내 성인 1,034명) 진행된 조사 결과에서는 '1년 안에 코로나19 백신이 개발된다면, 다음 중 어떤 결정을 내리시겠습니까?'라는 질문에 '안전성이 입증될 때까지 접종을 미룰 것이다'는 전체의 60.1%, '결코 접종을 받지 않을 것이다'는

1.4%, '아직은 모르겠다'는 3.5%의 응답으로, 백신 수용을 주저하거나 거부하는 의향이 전체의 65% 수준에서 나타났다고 볼 수 있다(그림 5-2-3).

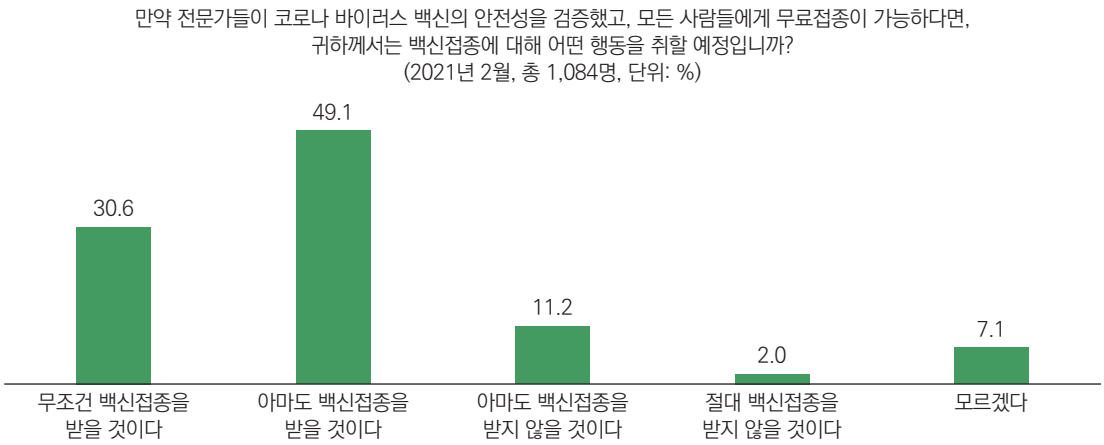


그림 5-2-4. 코로나19 백신 수용 의향(2021년)

한편, 2021년 1월(만 18세 이상 국내 성인 1,016명) 조사에서는 '코로나바이러스 백신이 출시된다면, 접종하고 싶습니까?'라는 질문에 전체의 37.5%가 맞을지 말지 '반반', 10.8%가 '아마 받지 않을 것', 4.9%가 '확실히 받지 않을 것'으로 응답한 것으로 나타났다. 그러나, 접종 사업 직전인 2021년 2월(만 18세 이상 국내 성인 1,084명) 조사에서는 '만약 전문가들이 코로나 바이러스 백신의 안전성을 검증했고, 모든 사람들에게 무료접종이 가능하다면, 귀하께서는 백신접종에 대해 어떤 행동을 취할 예정입니까?'라는 질문에 '아마도 백신접종을 받지 않을 것이다' 11.2%, '절대 백신접종을 받지 않을 것이다' 2.0%, '모르겠다' 7.1%로, 지연이나 거부의 의향이 낮은 수준을 보였다(그림 5-2-4).

2) 접종 사업 시기

2021년 11월 조사는 코로나19 백신을 접종했는가의 질문에 그렇다고 대답한 응답자를 대상으로, '추가접종(부스터샷)이나 지속접종(내년에도 접종)' 의향을 질문했다. 그 결과 '아마도 접종하지 않을 것이다'는 8.5%, '절대 접종하지 않을 것이다'는 3.1%, '아직 입장이 없다/잘 모르겠다' 9.5%로 백신 수용 의향이 낮거나 없는 비율이 높지 않았다. 그러나 한 해 뒤인 2022년 11월(만 18세 이상 성인 1,000명) 조사에서는 '추가 백신접종 권고가 있을 때의 접종의향'에서 '아마도 접종하지 않을 것이다' 22.8%, '절대 접종하지 않을 것이다' 5.0%, '아직 입장이 없다/잘 모르겠다'는 6.2% 등 전체의 30% 수준이 백신 수용 의향이 낮거나 없는 것으로 나타났다(그림 5-2-5).

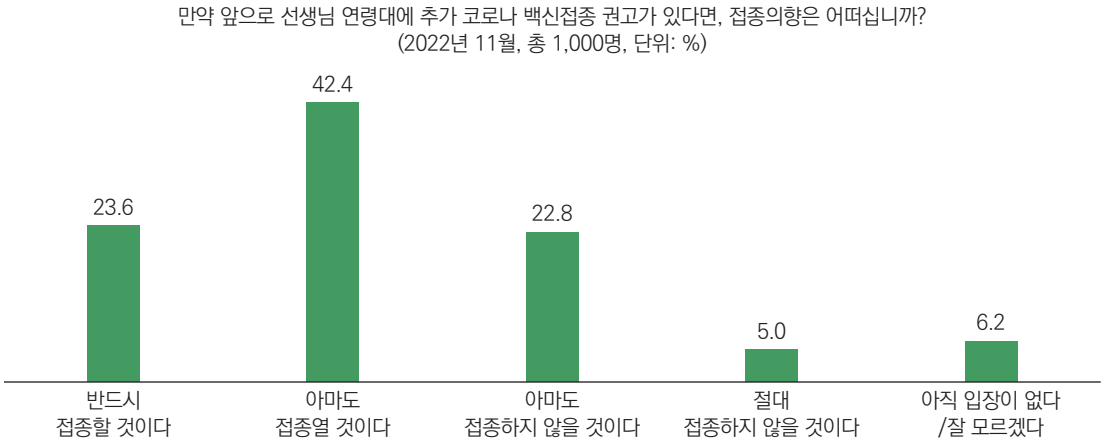


그림 5-2-5. 코로나19 백신 수용 의향(2022년)

백신 주저는 학부모 등 특정 집단 대상의 코로나19 예방접종 의향 조사에서 더 뚜렷하게 나타난다. 2022년 2월 초등학교 1~6학년 아동의 부모 113,450명을 대상으로 시행한 조사 결과, '접종 의향 있으나 좀 지켜보고 싶음'은 28.9%, '접종 의향 없으나 좀 지켜보고 싶음'은 22.9%, '접종 의향 없으며 아이에게 접종시키지 않을 것'은 38.1%, '모르겠음/확실하지 않음'은 3.6%로, 합산하면 전체의 93.5%가 백신 주저 의향을 보였다(그림 5-2-6).

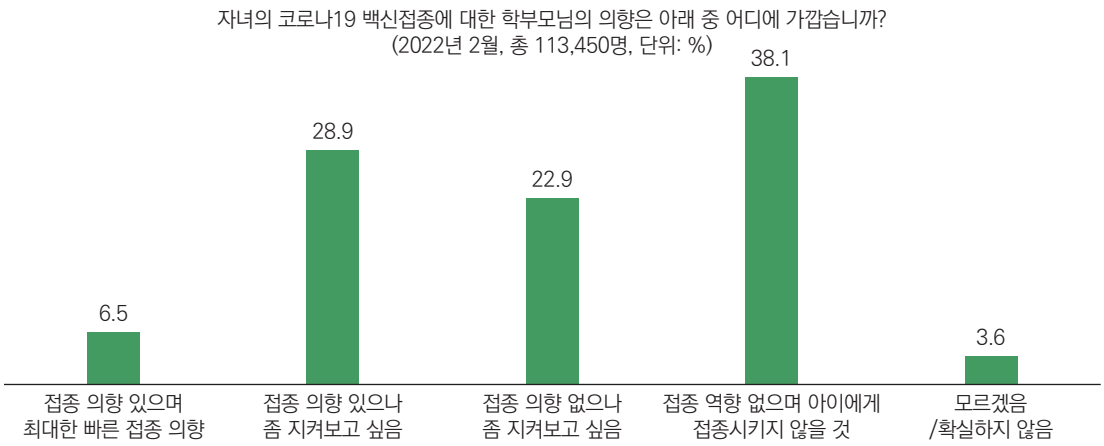


그림 5-2-6. 학부모의 코로나19 백신 수용 의향(2022년)

3) 연례 접종 시기

2023년 10월, 18세 이상 성인 1,000명을 대상으로 한 조사에서 '이번 동절기 코로나19 예방접종 의

향'을 질문한 결과 백신을 수용하지 않으려는 의향은 전체의 46.9%로 접종을 하겠다는 의향과 거의 유사한 수준을 보였다(그림 5-2-7).

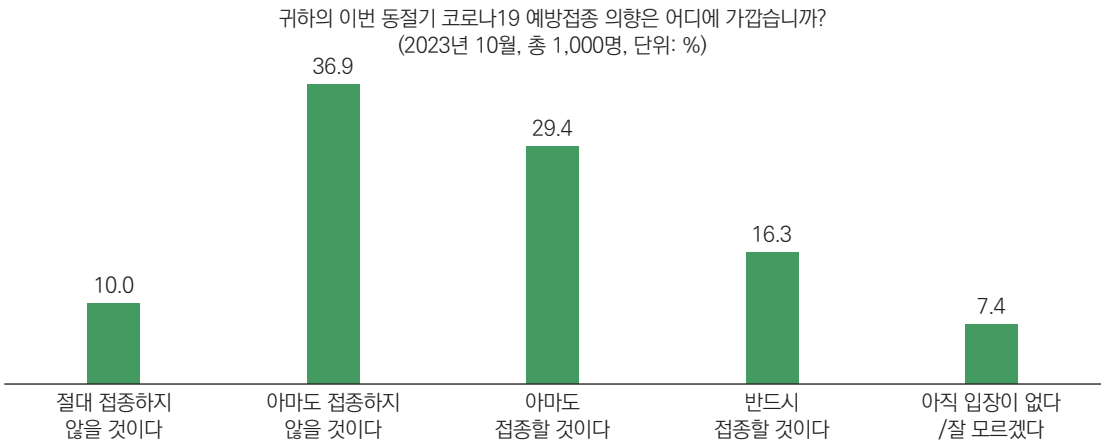


그림 5-2-7. 코로나19 백신 수용 의향(2023년)

4) 해외 조사 결과

Enea 등(2023)이 2020년 3~4월 사이 20개국 총 6,697명에게 "코로나바이러스 백신이 개발되면 접종할 의향이 얼마나 있습니까?"라는 질문을 한 결과 전체의 16.8%가 '결정하지 못했다', 4.4%는 '낮다', 6.0%는 '매우 낮다'고 응답하여 백신 수용 의향이 낮거나 부정적인 경우는 전체의 26% 수준인 것으로 나타났다. 또한, Wong 등(2021)이 2021년 1월부터 3월까지 17개국의 19,714명 대상으로 코로나19 예방 접종 의향을 조사한 연구에 따르면 접종 의향이 '높다' 또는 '매우 높다'인 긍정 의향은 호주(96.4%), 중국(95.3%), 노르웨이(95.3%) 순으로 높고, 반대로 낮은 접종 의향은 일본(34.6%), 미국(29.4%), 이란(27.9%)에서 상대적으로 높다. 이 중 미국은 '매우 낮다'의 응답 비율이 가장 높은 국가로 나타났다. 이 시기 조사들에 따르면 한국은 백신 수용 의향이 80%를 넘어선 국가의 하나로 꼽힌다.

5. 백신 주저의 영향 요인

현재 백신 주저 모델은 다수 존재한다. 그중 세 개를 예시하면 아래와 같다.

1) 세계보건기구의 BeSD 프레임워크

세계보건기구는 예방접종을 설명하는 행태적, 사회적 요인을 BeSD 프레임워크라고 이름 붙여 제

시하였다(그림 5-2-8). 이 모델은 개인의 정서와 인지의 요인(thinks and feels)과 주변의 영향(social process)이 개인의 동기에 영향을 주고, 그렇게 형성된 긍·부정적 동기가 예방접종의 참여 또는 거부로 이어지며 그 과정에 예방접종 사업의 실무적 요인(접근성, 비용과 부담 등)이 영향을 미친다고 가정한다.

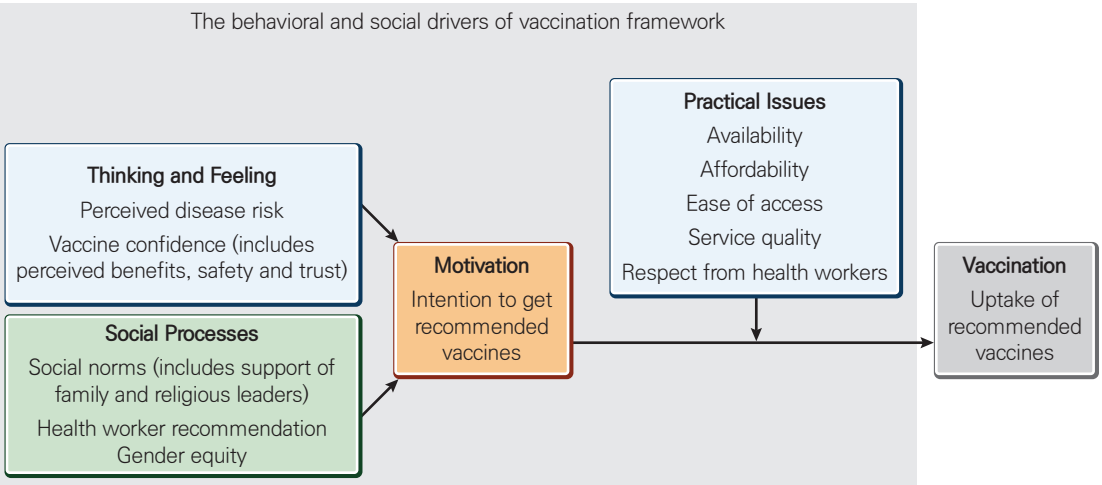


그림 5-2-8. 세계보건기구의 BeSD 프레임워크

2) Betsch 등의 5C 모델

Betsch 등의 5C 모델(The 5C Model of Vaccine Hesitancy)은 현재 세계보건기구 등을 통해서도 소개되고 있다. 이 모델은 초기에는 예방접종에 대한 신뢰와 확신(Confidence), 위험 판단의 현상 유지(Complacency), 접종의 편의(Convenience) 등 모두 알파벳 C로 시작하는 세 개의 요인으로 소개되었으나, 이후 이득-위험 계산(Calculation)과 공동체 보호와 같은 집단의 책임(Collective Responsibility)을 추가하여 현재의 5C 모델이 되었다(그림 5-2-9). 각 요인의 의미는 다음과 같다.

① Confidence(신뢰와 확신): 신뢰는 세 가지 수준을 포함한다. 백신의 효과성, 안전성에 대한 신뢰(백신 자체에 대한 신뢰), 예방접종을 권고하는 주체나 전문가에 대한 신뢰, 관련 시스템과 정책에 대한 신뢰이다. 신뢰가 낮다면, 백신 불신이나 거부로 이어질 수 있다.

② Complacency(현상 유지): 예방접종을 우선순위에서 낮게 보는 것으로, 백신으로 예방 가능한 질병에 대한 인지된 위험도를 반영한다. 만약 해당 감염병을 심각한 위협으로 보지 않는다면, 예방접종 없이 현상을 유지하려는 자족적 태도나 의향이 촉발된다.

③ Convenience(편의): 예방접종의 물리적·경제적·심리적 접근성과 관련된 것으로 불편, 장벽, 제

약을 느끼는 정도이다. 이는 Constraints(제약/제한)이라고도 한다. 예방접종에 심적 부담이 크고, 경제적 유인이 없고, 물리적으로 불편하거나 제약이 있다면, 예방접종을 주저하거나 기피로 이어진다.

④ Calculation(이득-위험 계산): 감염의 위험과 예방접종의 이득을 가늠하는 것이다. 위험과 이득을 타산하기 위한 정보를 수집하고 탐색하는 노력 및 참여와 관련이 있다.

⑤ Collective Responsibility(공동체 보호/집단 책임): 예방접종을 통해 타인과 공동체를 보호하겠다는 적극적 의지를 말한다. 높은 집단 책임성과 공동체 보호 의지는 예방접종에 대한 긍정적 의향과 실제 접종을 높이는 방향으로 기여할 수 있다.

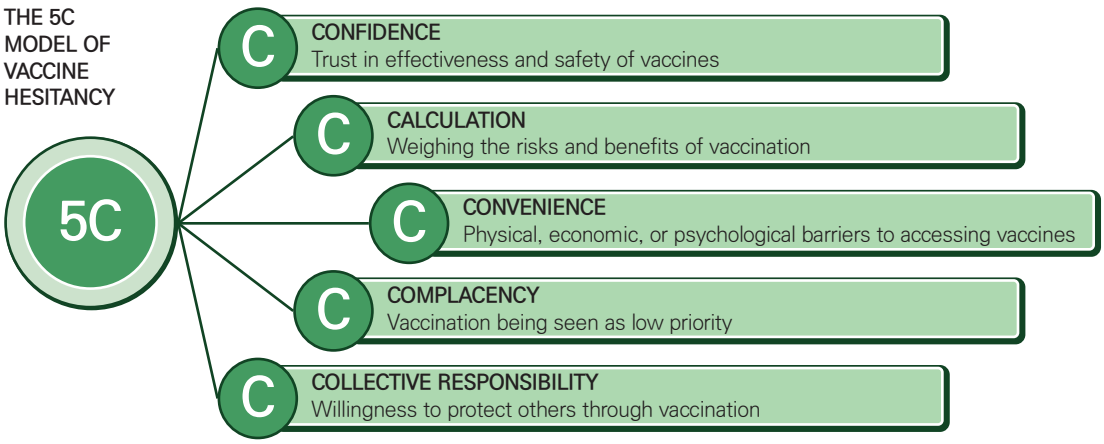


그림 5-2-9. Betsch 등의 5C Mode

5C 모델을 적용한 실증 연구들은 백신 수용 의향에 5C가 고루 유의미한 영향을 미치고, 그중에서도 Confidence의 높은 기여를 보고하고 있다. 예를 들어 Wismans 등(2021)은 신뢰와 공동체 보호 동기가 학생들의 코로나19 예방접종 의도를 설명하는 가장 중요한 요인으로 보고했으며, Rancher 등(2023) 역시 이 두 요인의 유의미한 영향력을 보고했다.

3) Dubé 등의 백신 주저에 대한 개념적 틀

Dubé 등의 백신 주저 개념 모델은 5C 모델보다 거시적인 개념 틀이다. 이 모델은 예방접종에 대한 의사결정이 단순히 개인적 요인에 의해서만 이루어지는 것이 아니라, 역사·정치·사회·문화적 맥락 속에서 다양한 요인과 신뢰가 결합하여 복합적으로 형성된다는 점을 강조한다(그림 5-2-10). 각 요인의 간단한 설명은 아래와 같다.

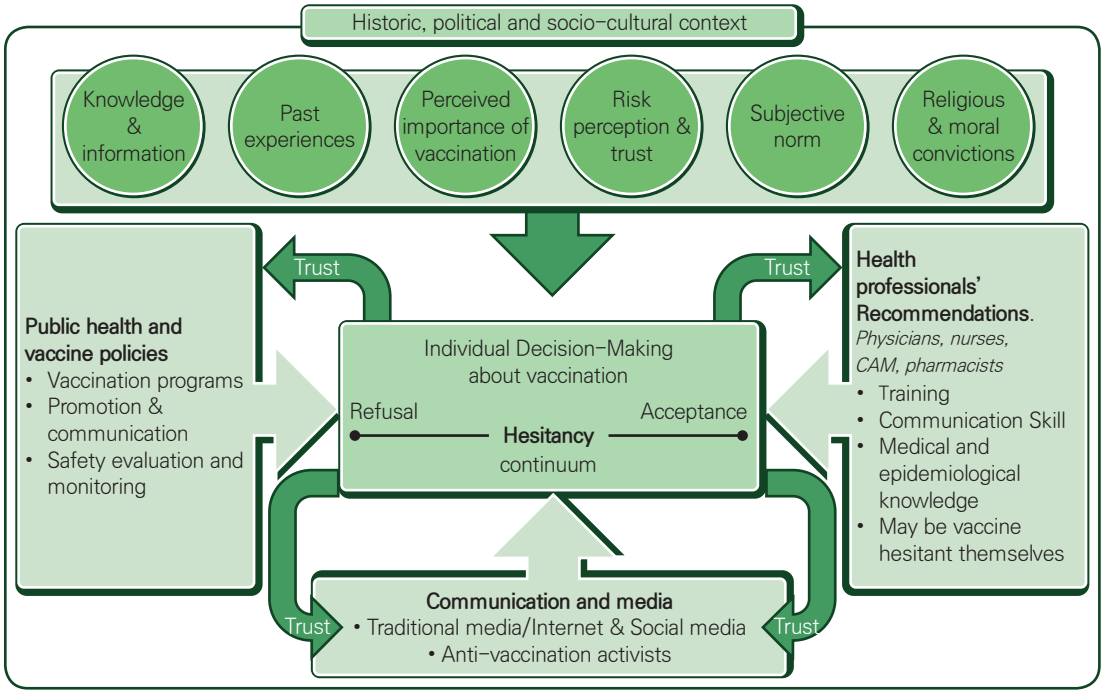


그림 5-2-10. 백신 주저의 개념 모델

(1) 역사, 정치, 사회문화의 맥락(Historic, Political, and Socio-Cultural Context)

- 지식과 정보(Knowledge & Information): 백신에 대한 이해와 지식 수준
- 과거 경험(Past Experiences): 개인의 과거 접종이나 백신 관련 경험
- 백신의 중요성 인식(Perceived Importance of Vaccination): 예방접종이 질병 예방에 얼마나 중요한지에 대한 개인의 인식
- 위험 인식과 신뢰(Risk perception & Trust): 백신의 안전성과 효과에 대한 신뢰 및 인식을 의미
- 종교적 및 도덕적 신념(Religious & Moral Convictions): 종교적 가치와 도덕적 기준이 백신 결정에 미치는 영향력

(2) 공중보건 및 백신 정책(Public Health&Vaccine Policies)

예방접종 사업 설계, 추진, 홍보와 소통을 통한 대중 교육, 백신 안전성 평가 및 모니터링 등 정부와 보건 기관의 역할

(3) 의료 전문가의 권고(Health Professionals' Recommendations)

의사, 간호사, 약사 등 의료 제공자가 훈련, 의사소통 기술, 지식 제공 등을 통해 비전문가인 개인의 예방접종 의향과 결정에 미치는 영향력(예: 의료 전문가가 예방접종을 주저할 수 있음)

(4) 소통과 미디어(Communication and Media)

전통 미디어, 소셜 미디어 등 채널을 통해 백신 주저를 조장할 수 있음

(5) 신뢰(Trust)

신뢰는 모든 요인과 개인 의사결정 과정 사이를 연결하는 핵심 요소로, 공중 보건 정책, 의료 전문가, 그리고 커뮤니케이션 채널에서 신뢰가 강화되면 백신 수용 가능성이 높아질 수 있음

(6) 개인의 예방접종 의사결정 과정(Individual Decision-Making about Vaccination)

백신 주저를 거부(refusal)에서 수용(acceptance)으로 이어지는 연속선(continuum)으로 나타내며, 이러한 결정은 다양한 외부 요인과 신뢰(trust)의 영향을 받는다.

3 백신 주저의 대응 방안

앞으로도 공중보건의 목표를 달성하기 위해서는 예방접종을 통한 개인과 집단의 면역 획득을 강조하게 될 것이다. 그렇다면 관건은 공중의 백신 신뢰와 수용을 높일 더 나은 방안을 마련하는 것이다. 소통의 측면에서 어떤 대안이 가능할지에 대해 국내외 보건기구와 문헌이 권고하는 내용을 요약하면 아래와 같다.

1. 유형별 대응 방안

백신 소통의 효과성 향상을 위해 '맞춤(tailoring)'을 강화할 필요가 있다. 세계보건기구는 과거 팬데믹 및 감염병 백신(H1N1, 소아마비 등)과 관련된 경험과 연구를 바탕으로, 코로나19 예방접종 의향의 유형을 제시했다(그림 5-2-11). 주의가 필요한 것은 이들 유형은 비례적인 것이 아니라서 상대적 비율이나, 소통 대응의 우선순위 등은 반영되지 않은 단순 도식이란 점이다.

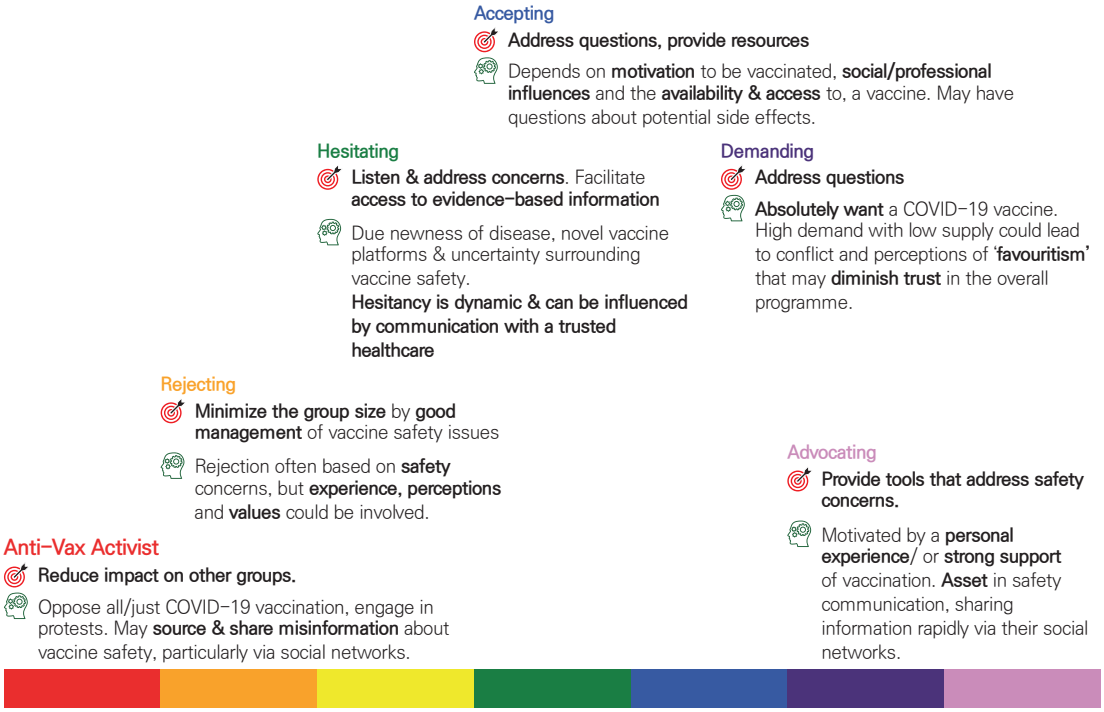


그림 5-2-11. 코로나19 백신 의향 스펙트럼
(Spectrum of vaccination intentions for COVID-19 vaccines)

1) '안티 백서'와 백신 거부 유형 대응

'안티 백서(Anti-Vax Activist)' 활동가 또는 백신 거부(Rejecting) 집단은 모든 백신을 거부하고 백신의 거부와 반대를 개인 수준이 아닌 집단 수준으로 조직하는 활동가들이다. 세계보건기구는 이 집단이 다른 집단에 미치는 영향력을 줄이는 것을 말하지만 강한 백신의 반대와 거부에는 백신 안전성에 대한 부정적 경험이 뿌리에 있을 수 있고, 또한 반대를 위한 반대가 아닌 정확하고 투명한 정보를 제공하는 것을 원칙으로 이 유형에 대응할 필요가 있다고 지적한다.

2) 백신 주저 유형 대응

백신 주저(Hesitating) 집단은 해당 질병의 새로움과 낯섦, 새로운 백신 플랫폼의 사용, 백신 안전성에 대한 불확실성 등을 종합적으로 우려하고, 그 결과 예방접종을 꺼리거나 망설이게 된다고 이해해 볼 수 있다. 이러한 태도는 예방접종 프로그램에 익숙해지면서 변할 수 있는데, 이를 위해 백신 안전성에 대한 우려를 경청하고 공감하며, 전문가와 소통하여 의구심을 해소할 수 있는 기회를 여는 것이 중요하다. 무엇보다도 신뢰할 수 있는 증거 기반 백신과 예방접종 정보를 널리 제공하여, 주저나 기피 의향을

가진 공중이 적절한 정보를 바탕으로 현명한 결정을 내릴 수 있도록 지원하고 돕는 접근을 높여야 한다.

3) 백신 수용(Accepting) 유형 대응

이 집단에 속하는 사람의 대부분은 예방접종에 긍정적인 태도를 보이는데, 일부는 백신을 강력히 요구한다(Demanding). 또한 백신을 옹호하고 지지(Advocating)하는 사람들은 자신들의 소셜 네트워크를 통해 백신 소통에 유의미한 역할을 할 수 있다. 그러나, 예방접종을 옹호하거나 긍정적으로 수용하더라도, 이들 모두는 백신과 예방접종의 잠재적인 위험과 부작용에 의문을 제기할 수 있다는 점을 신중히 고려해야 한다. 따라서, 우려와 의문에 성실히 응답하고, 투명하게 다루어서 건설적인 백신 지지와 옹호의 네트워킹을 촉진하도록 노력한다.

그 외 미국 질병통제예방센터(CDC)는 지역사회의 예방접종 수준을 근거로 유형을 구분하고 접근하는 방안을 제시했다.

① 데이터 기반으로 접종 참여의 가능성이 낮은 지역사회를 파악하고, 우선순위를 정한다.
② 해당 지역의 공중보건, 민간, 기타 영역에서 주요 백신 소통과 사업의 파트너를 찾는다.
③ 해당 지역사회에 가로놓인 예방접종의 물리적 사회심리적 장애물과 어려움을 파악하고 대안을 위한 계획을 수립한다.
④ 위 ②번에서 구축한 지역사회 파트너십 네트워크가 예방접종 사업의 기획과 실행 전 단계에 참여하고 함께 일하도록 실질적인 방안을 마련한다.
⑤ 해당 지역사회의 예방접종 사업을 평가하고 변화가 필요한 지점들을 개선하여 반영한다.

2. 소통 중심의 대응 방안

백신 소통은 기본적으로 백신이 안전하게 개발되었고, 예방접종이 효과적임을 알리는 노력이라고 볼 수 있다. 그러나 코로나19 백신 소통은 이것으로 충분하지 않다는 교훈을 남겼다. 향후 더 효과적인 백신 소통을 위해 참고할 내용은 아래와 같다.

1) 백신 소통의 핵심 원칙

첫째, 투명성과 개방성을 최우선으로 한다. 정보를 있는 그대로 모두 접근할 수 있게 하고, 정보 공유를 반드시 준수한다. 정기적으로 업데이트하고 부족한 점은 인정하며 보완한다.

둘째, 반응성을 높인다. 확실해질 때까지 기다리기보다 현재의 제한점을 분명히 하면서 최신 정보를 신속히 제공한다. 정부 조치를 업데이트하고 백신이나 예방접종 관련 문제 발생 시 즉시 공유한다. 언론이나 소셜 미디어와 협력하여 간결한 메시지를 실시간으로 전달한다.

셋째, 경청과 공감을 강조한다. 백신을 둘러싼 긍·부정적 입장과 감정을 인정하고 경청한다. 다양한 의견에 대해 논쟁하거나 바로잡으려 하지 않으며, 개방적 태도를 지속한다.

넷째, 불확실성을 효과적으로 다룬다. 단순화하거나 과도한 확신을 전달하려고 하지 않는다. 공중이 숙고할 내용을 시나리오별로 제시하고 필요한 조치를 할 수 있는 것과 불확실성이 있는 것을 명확히 설명한다.

다섯째, 공중의 건강 문해력을 고려한 자료와 성명을 개발하고, 쉬운 언어로 명확하게 메시지를 전달하며 청중의 요구에 적극적으로 대응한다.

2) 소통 단계별 주요 내용

(1) 소통 계획 단계

예방접종이 시작되기 전 단계의 핵심은 정부당국이 백신 소통을 위한 효과적인 구조를 마련하는 것이다. 소통 창구와 담당자를 지정하는 등 과업을 나누고, 핵심 소통 집단을 선정하고 소통의 우선순위를 설정한다. 해당 집단의 문해력(헬스 리터러시), 온라인 정보 접근성, 정보통신 기술 활용 수준 등을 세심히 파악하고, 그에 기초하여 소통의 효과를 높일 세부 계획을 수립한다.

동시에 대외적으로 백신 소통의 협력 구조를 확보하는 것이 중요하다. 대외적으로 신뢰가 높은 전문가, 소셜 미디어의 영향력이 높은 인플루언서, 그 외 백신 소통의 다양한 중요 스피커를 확보하고 상호 협력의 동기와 방안을 사전에 확보한다. 특히, 미디어와의 사전 협력 체계의 마련이 중요하다.

(2) 소통 실행 단계

백신의 안전성, 예방접종 효과성을 소통하는 실제 단계에서는 사전에 확립한 가이드라인에 따라 차질 없이 소통한다. 이 단계에서 가장 중요한 것은 청취와 의견수렴이다. 소통은 정보 전달만이 아니라 듣고 반영하는 피드백의 노력이 핵심임을 잊지 않아야 한다. 청취의 방법은 설문이나 면접, 집중 자문과 소셜 미디어 데이터를 활용한 공중의 실시간 의견 모니터링까지 다양하다. 이 과정에서 부정적 의견의 내용을 탐지하고, 주변에 미치는 영향력을 주의 깊게 살핀다.

예방접종에는 다양한 정서적 반응이 있다고 가정하고, 이해와 공감을 표시하며 주요 정보와 소통 수요를 정리해 내는 역량이 필요하다. 같은 맥락에서, 언론의 정보 요구에 효과적으로 반응해야 한다. 백브리핑, 취재 지원 등을 통해 공중이 미디어를 통해 정확하게 중요한 정보를 얻고 활용하도록 집중 지

원한다.

이 소통 실행 단계에서 중요한 또 하나의 과제는 백신 소통 속에서 형성되는 의미를 파악하는 것이다. 만일 매우 부정적인 서사나 사례가 널리 공유되어, 그로부터 공중의 백신 신뢰가 위협을 받는다면, 그 대응을 효과적으로 조직해야 한다. 신뢰받는 스피커(전문가 등)들과 사전에 구축한 협력 체계를 통해 전문가의 백신 소통이 진정성을 얻고 설득력을 얻도록 지원하는 것도 이 단계의 중요 활동이다.

(3) 소통 피드백(평가) 단계

백신 사업이 종료되면 백신 소통을 반드시 평가하여 문제의 원인을 찾고 개선점을 도출하는 단계를 밟는다. 구체적으로, 소통의 효과 평가, 도전 과제와 교훈 기록, 기술 및 자원 격차 확인, 개선 방안 도출이 포함된다. 이를 위해 소셜 미디어 청취, 미디어 모니터링, 의료인 등을 통하거나 지역사회 기반(지역과 종교계 지도자, 시민사회단체 등)으로 피드백을 구할 수 있다.

3) 소통 요소별 대응 방안

(1) 미디어 대응

전통 미디어와 소셜 미디어 모두 백신 정보를 전달하는 중요 매체로, TV와 라디오 등 전통 미디어는 노인층이 예방접종 정보를 얻는 주요 출처이다. 따라서 전통 미디어에서는 메시지-수용자 맞춤 수준을 높일 필요가 있다. 예컨대, 예방접종으로 중증 진행을 예방하는 효과가 있음을 간결한 메시지로 전할 수 있도록 한다. 고령층에게 신뢰가 높은 의료 전문가를 통해 과학의 불확실성을 효과적으로 다루는 것도 중요하다. 특히, 고령층의 과학 문해력을 높이는 장기적인 노력도 필요하다. 소셜 미디어의 경우 실시간 양방향 소통이 가능한 장점이 있지만 대신 별도의 전담팀이 필요하다. 만일 소셜미디어 대응 전담 구조가 어렵다면, 모든 소셜 미디어 대응보다 주요 플랫폼에 가용 자원을 집중하는 전략이 효과적이다. 이때 해당 플랫폼에서 영향력이 높은 인플루언서와 협업을 모색하여 지렛대 효과를 얻는 것이 유익하다. 반대로 부정확하고 왜곡된 정보가 집중되는 플랫폼을 파악하고 그 플랫폼에 가장 효과적으로 정확하고 객관적인 정보를 제공할 전략을 찾고 적용해야 한다.

(2) 백신 소통 메시지

객관적이라도 이해가 어렵거나 활용하기 힘들다면 무용지물이라는 인식이 필요하다. 지나치게 복잡한 전문 용어를 최소화하고, '쉬운 언어(plain languages)' 전략 등을 활용하여 수용자가 편안함을 느끼도록 백신 소통 메시지를 정제할 필요가 있다. 또한, 건조한 것이 과학 정보라는 인식을 탈피하고 설득력 있는 서사(narrative)로 예방접종의 기대효과를 공유할 수 있도록 전문성을 발휘할 필요가 있다.

이때 수용자나 청중의 대표자와 백신 소통 콘텐츠를 사전에 점검하는 것이 매우 중요하다. 그것만이 공중의 반응을 예측하기 어려운 상황에서도 혼란을 막고 신뢰를 확보할 수 있다는 것을 원칙으로 분명히 해야 한다. 자주 묻는 질문(FAQs) 등을 활용하여 소통 효율성을 높일 수 있는데, 이 또한 사전 점검과 피드백을 거쳐야 한다.

4) 그 외 방안

(1) 성별이나 연령대 맞춤

현재 문헌들은 여성이 남성보다 백신 수용 의향이 전반적으로 낮은 경향성을 보인다고 지적하고 있다. 이를 감안하여, 백신의 안전성을 다루면서 임신 및 수유 중 접종이 안전하다는 것을 효과적으로 전달할 방안을 마련할 필요가 있다. Syan 등(2021)은 전문가의 직접 설명 기회를 늘리고, 여성의 고유한 우려나 의문점에 맞춤형 답변을 개발하여 제공할 필요성을 제시했다. 남성의 경우는 감염 위험을 상대적으로 낮게 보는 인식에 대응할 전략이 필요하다. 또한, 예방접종의 공동체 보호 동기를 고취할 소통 전략도 권고된다.

Afifi 등(2021)은 청소년에게는 시각 자료와 교육 프로그램을 통한 백신 소통을, 청년층에게는 강요가 아닌 자율적 의사결정을 강조하는 백신 소통이 더 효과적이라고 권고한다. 또한, 두 집단 모두 신뢰하는 정보 출처가 부모 등보다 또래 집단과 SNS 중심으로 이동하는 특징을 보인다는 점을 감안할 필요가 있다는 지적도 참고할 만하다. 고령자의 경우, 하위 집단에 따라 고유한 인식이 있음을 유념할 필요가 있다.

Silvonen 등(2023)의 질적 연구 등을 종합하면 'young-old' 집단(대략 55~64세)은 자신의 건강 상태가 양호하다고 판단하여 "건강에 이상이 없으니 예방접종은 불필요하다"는 인식을 나타낼 수 있고, 65~74세 집단은 상대적으로 예방접종 의향이 높은 편이지만 여러 개의 백신을 접종하는 것에 대해 의사결정을 내리기 위해 정보가 부족하다는 인식을, 그 이상의 연령대(old-old)는 백신을 수용하는 경향이 강하고 지난 예방접종 경험과 의료 전문가의 영향력이 높다고 보고되었다. 특히, 가족이나 보호자의 부정적 의견이 당사자의 예방접종 의향에 미치는 영향력이 높다고 보고되는 것을 고려하여 가족 단위의 백신 소통 방안도 고려할 필요가 있다. 이때 중요한 것이 지역사회의 참여이다. 해당 지역 고령층을 누구보다 잘 아는 공중보건, 사회복지, 미디어, 기타 지역의 주요 리더와 영향력 집단이 고령층 건강 보호를 위한 백신 소통에서 협력하는 것이 중요하다.

5) 결론

과거 공중보건은 개인이 자신과 공동체의 건강을 위해 중요한 정보에 접근하고, 그 정보를 활용하여

합리적인 판단을 내린다고 가정했다. 이때 습득한 건강 지식과 정보는 긍정적 의향과 동기를 촉진하고, 그 결과 건강증진 행태가 일어난다고 본 것이다. 그러나 현실은 위기나 위험의 상황에서 개인의 지식과 인식과 태도와 행태가 단선적이지 않다는 것을 보여주었다. 오히려 개인의 의향과 의사결정 사이에는 지식이나 정보의 양과 수준 이상의 많은, 사회심리적, 문화적 제도적 요인이 영향을 미친다는 것을 보여주었다. 바로 이 과정에서 정보와 의견의 소통이 차지하는 영향력이 점차 그 중요성을 높여가고 있다.

전대미문의 코로나19 팬데믹을 맞이하여 보건당국은 과학 전문가, 의료 제공자, 미디어 그리고 지역사회 공중과 함께 그 위기에 맞섰다. 이 과정에서 안전성과 효과성이 입증된 백신의 접종이 적극 권고되었다. 실제 우리나라 공중은 적극적인 예방접종 참여를 통해 지금의 일상 회복을 이끌었다. 문제는 코로나19 같은 공중보건 위기의 가능성은 여전히 상존한다는 것이다. 이는 일상의 맥락에서 감염 위험 관리 역량 강화가 필요함을 의미한다.

각자의 역할이 있다. 과학은 안전하고 효과적인 백신과 치료제 개발에 박차를 가하고 정부와 공중보건 체계는 예방접종을 포함하여 감염병 관리 사업의 신뢰와 효과를 높이기 위한 전력을 다해야 한다. 무엇보다도 공중이 백신에 대한 과학적 이해도를 높이고, 부정적이거나 부정확한 정보로부터 스스로와 주변을 보호하고, 자율적인 예방접종 의사결정을 내리도록 돕고 지원해야 한다. 백신 소통은 이 가운데에 위치하면서 모두를 관통한다. 본 절이 제시한 내용을 바탕으로 우리나라 공중보건 분야의 핵심 인력인 역학조사관의 백신 소통 역량 강화가 이뤄지는 것이 대단히 중요하다.

참고 문헌

1. 서울대학교 보건대학원 보건조직과 건강커뮤니케이션 연구실. 동절기 코로나19 대국민 인식조사(2022.11.03~11.07, 전국 만18세 이상 1,000명). 2022.
2. 서울대학교 보건대학원 보건조직과 건강커뮤니케이션 연구실. 코로나19 국민인식조사 7차(2020.10.27~10.29, 전국 만 18세 이상 1,000명). 2020.
3. 서울대학교 보건대학원 보건조직과 건강커뮤니케이션 연구실. 코로나19 대응 체제 전환에 관한 국민 인식조사(2021.11.22~11.29, 전국 만 18세 이상 912명). 2021.
4. 서울대학교 보건대학원 보건조직과 건강커뮤니케이션 연구실. 코로나19와 사회적 건강 2차 조사(2021.02.08~02.17, 전국 만18세 이상 1,084명). 2021.
5. 서울대학교 보건대학원 보건조직과 건강커뮤니케이션 연구실. 한국리서치 정기조사(2023년 19차_10월 2주). 2023.
6. 질병관리청. 2024년도 예방접종대상 감염병 관리지침. 2024.
7. 질병관리청. 코로나19 국내 발생 및 예방접종 현황 (10월 28일). 2022.
8. Adhikari B, Cheah PY, von Seidlein L. Trust is the common denominator for COVID-19 vaccine acceptance: a literature review. Vaccine: X. 2022;12:100213.
9. Afifi TO, Salmon S, Taillieu T, et al. Older adolescents and young adults willingness to receive the COVID-19 vac-

cine: Implications for informing public health strategies. Vaccine.2021;39(26):3473-9.

10. Bavel JJV, Baicker K, Boggio PS, et al. Using social and behavioural science to support COVID-19 pandemic response. Nature Human Behaviour. 2020;4(5):460-71.
11. Betsch C, Schmid P, Heinemeier D, et al. Beyond confidence: Development of a measure assessing the 5C psychological antecedents of vaccination. PLoS One. 2018;13(12):e0208601.
12. Caron RM, Noel K, Reed RN, et al. Health promotion, health protection, and disease prevention: Challenges and opportunities in a dynamic landscape. AJPM Focus. 2023;3(1):100167.
13. Centers for Disease Control and Prevention (CDC) [Internet]. Partnering for vaccine equity. Atlanta: CDC [Cited 2021 Oct 26]. Available from: https://www.cdc.gov/vaccine-equity/php/about/index.html.
14. Communication theory and its applications in nursing and healthcare [Internet]. Nurse Key [Cited 2025 Jan 17]. Available from: https://nursekey.com/communication-theory-and-its-applications-in-nursing-and-healthcare/.
15. Enea V, Eisenbeck N, Carreno DF, et al. Intentions to be vaccinated against COVID-19: The role of prosociality and conspiracy beliefs across 20 countries. Health Communication. 2023;38(8):1530-9.
16. Gallotti R, Valle F, Castaldo N, et al. Assessing the risks of 'infodemics' in response to COVID-19 epidemics. Nature Human Behaviour. 2020;4(12):1285-93.
17. Habersaat KB, Jackson C. Understanding vaccine acceptance and demand—and ways to increase them. Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz. 2020;63(1):32-9.
18. Lazarus JV, Ratzan SC, Palayew A, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nature Medicine. 2021;27(2):225-8.
19. Lazarus JV, Romero D, Kopka CJ, et al. A multinational Delphi consensus to end the COVID-19 public health threat. Nature. 2022;611(7935):332-45.
20. Lee M, Seo S, Choi S, et al. Parental acceptance of COVID-19 vaccination for children and its association with information sufficiency and credibility in South Korea. JAMA Network Open. 2022;5(12):e2246624.4.
21. Lee M, You M. Direct and indirect associations of media use with COVID-19 vaccine hesitancy in South Korea: cross-sectional web-based survey. Journal of Medical Internet Research. 2022;24(1):e32329.
22. MacDonald N, Dubé È. Canadian guidance on addressing vaccine hesitancy to help foster vaccine demand and acceptance. Ottawa (ON): CANVax; 2019.
23. Mohseni Afshar Z, Barary M, Hosseinzadeh R, et al. COVID-19 vaccination challenges: A mini-review. Human Vaccines&Immunotherapeutics. 2022;18(5):2066425.
24. Nwachukwu G, Rihan A, Nwachukwu E, et al. Understanding COVID-19 vaccine hesi- tancy in the United States: a systematic review. Vaccines. 2024;12(7):747.
25. Porat T, Nyrup R, Calvo RA, et al. Public health and risk communication during COVID-19—enhancing psychological needs to promote sustainable behavior change. Frontiers in Public Health. 2020;8:573397.
26. Rancher C, Moreland AD, Smith DW, et al. Using the 5C model to understand COVID-19 vaccine hesitancy across a national and South Carolina sample. Journal of Psychiatric Research. 2023;160:180-6.
27. Randolph HE, Barreiro LB. Herd immunity: understanding COVID-19. Immunity. 2020;52(5):737-41.
28. Rimal RN, Lapinski MK. Why health communication is important in public health. Bulletin of the World Health Organization. 2009;87:247-a.
29. Sallam M. COVID-19 vaccine hesitancy worldwide: a concise systematic review of vaccine acceptance rates. Vaccines. 2021;9(2):160.
30. Silvonen T, Kesten J, Cabral C, et al. Views and perceptions of people aged 55+ on the vaccination programme for older adults in the UK: a qualitative study. Vaccines. 2023;11(4):870.

31. Su Z, McDonnell D, Wen J, et al. Mental health consequences of COVID-19 media coverage: The need for effective crisis communication practices. Globalization and Health. 2021;17:1-8.
32. Syan SK, Gohari MR, Levitt EE, et al. COVID-19 vaccine perceptions and differences by sex, age, and education in 1,367 community adults in Ontario. Frontiers in Public Health. 2021;9:719665.
33. Wismans A, Thurik R, Baptista R, et al. Psychological characteristics and the mediating role of the 5C model in explaining students' COVID-19 vaccination intention. PLoS One. 2021;16(8):e0255382.

Textbook For
Korea Field Epidemiology
Training Program





6장

의료관련감염과 의료관련감염병의 특성과 관리







제1절 의료관련감염과 의료관련감염병의 정의와 특성
제2절 감염예방 및 관리의 종류와 적용범위
제3절 의료관련감염병 감시체계
제4절 의료관련감염 및 의료관련감염병 유행 시 역학조사 단계와 방법

집필자 김진용, 유석현

제1절 의료관련감염과 의료관련감염병의 정의와 특성

미국 질병통제예방센터에 따르면 입원환자 31명 중 1명이 최소 1개 이상의 의료관련감염병에 매일 감염되며 이 중 6%는 사망한다고 보고하고 있다.

1 의료관련감염의 정의

의료관련감염(Healthcare-Associated Infection, HAI)이란 입원뿐만 아니라 외래진료를 포함하여 의료기관내에서 의료행위와 관련된 감염을 말하며, 이는 병원근무자 등 관련 종사자들의 감염까지 포함한다. 의료관련감염 발생은 환자군의 특성, 병원의 특성, 감염의 종류에 따라 다르지만, 입원환자의 5~10%에서 발생하는 것으로 추정되며, 특히 중환자실을 중심으로 요로감염, 혈류감염, 폐렴 순으로 발생빈도가 높은 것으로 알려져 있다. 의료관련감염병은 의료관련감염 중 「감염병의 예방 및 관리에 관한 법률」 제2조 제12호에 따라, 법적으로 관리되는 일부 감염병을 말한다.

표 6-1-1. 국내외 원내감염 및 의료관련감염병의 정의

	정의	종류
미국 질병통제센터 1988	원내감염(nosocomial infections)이란, 환자가 의료 기관에서 의료진의 어떠한 시술로 인하여 의료진이 관찰한 감염을 의미하며, 입원 후 잠복기 등을 고려한 기간 내에 새로이 발생하는, 지역사회 획득 감염 등이 제외된 감염병을 말한다.	수술 부위 감염, 혈류감염, 폐렴, 요로감염
세계보건기구 2012	의료관련감염병이란, 입원 당시에는 존재하지 않았거나 잠복하지 않았던 병원 균이 병원이나 기타 의료 시설에서 치료 과정 중 새롭게 환자에게 감염되는 것을 의미한다.	의료 관련 요로감염, 수술 부위 감염, 병원 내 폐렴, 인공호흡기 관련 폐렴, 의료 관련 혈류 감염

	정의	종류
질병관리청 2017	의료관련감염병이란, 입원 당시에 없었던 혹은 잠복하고 있지 않았던 감염이 입원기간 중 발생한 것으로 환자뿐 아니라 병원에서 발생하는 직원들의 감염도 이에 포함될 수 있으며, 퇴원 후에도 발생하는 경우도 있다.	삽입기구 관련 감염(인공호흡기 관련 폐렴, 중심정맥관 관련 감염, 유치 도뇨관 관련 감염)
유럽 질병예방 통제센터 2024	의료 관련 감염은 환자가 병원이나 의료 시설에 머무는 동안 발생하는 감염을 뜻한다.	가장 흔하게 보고되는 의료 관련 감염 유형은 호흡기 감염, 수술 부위 감염, 요로 감염, 혈류 감염 및 위장관 감염이며, 그 중 *Clostridium difficile* 감염이 위장관 감염의 절반을 차지한다.

2 국내 의료관련감염병의 정의 및 종류

우리나라에서 의료관련감염병이란, 환자나 임산부 등이 의료행위를 적용받는 과정에서 발생한 감염병으로서 감시활동이 필요하여 질병관리청장이 고시하는 감염병을 말한다. 「감염병의 예방 및 관리에 관한 법률」 제2조 제12호에 따른 의료관련감염병의 종류는 다음 각 목과 같다.

표 6-1-2. 국내 의료관련감염병의 종류

감시방법	감염병명	영문명
전수	반코마이신내성황색포도알균 감염증	Vancomycin-Resistant *Staphylococcus aureus* (VRSA) infection
	카바페넴내성장내세균목 감염증	Carbapenem-Resistant *Enterobacterales* (CRE) infection
표본	반코마이신내성장알균 감염증	Vancomycin-Resistant *Enterococci* (VRE) infection
	메티실린내성황색포도알균 감염증	Methicillin-Resistant *Staphylococcus aureus* (MRSA) infection
	다제내성녹농균 감염증	Multidrug-Resistant *Pseudomonas aeruginosa* (MRPA) infection
	다제내성아시네토박터바우마니균 감염증	Multidrug-Resistant *Acinetobacter Baumannii* (MRAB) infection

3 의료관련감염 및 의료관련감염병의 특성

법률로 지정된 의료관련감염병(주요 다제내성균)은 환자 치료에 대한 의료비 상승 및 사망률 증가 등을 야기할 수 있다. 다제내성균은 병원 내에서 환자 또는 병원체 보유자와의 직/간접적 접촉이나, 오염된 기구 등의 환경을 통하여 전파될 수 있으며, 면역력이 저하된 환자나 장기 입원환자에서 특히 위험하다고 알려져 있다.

의료관련감염 및 의료관련감염병은 다른 유형의 감염과 구별되는 몇 가지 주요한 특징이 있다. 첫째, 의료 환경에서 세균 및 바이러스 등의 광범위한 병원체의 감염으로 인해 발생할 수 있다. 대표적인 병원체로는 메티실린내성황색포도상구균(MRSA), 녹농균(Pseudomonas) 등이 있으며, 다제내성으로 인한 재원일수 등의 의료비용 부담이 증가하게 된다. 둘째, 의료관련감염병의 의료비용 부담을 증가시키는 요인은 환자 내부적 요인과 외부적 요인 등 다양할 수 있다. 환자 내부적 요인으로는 환자가 갖고 있는 기저질환이나 약화된 면역체계가 있을 수 있으며, 환자 치료에 따른 입원 기간(장기간 또는 빈번한 입원)이 해당한다. 정상적인 미생물 세균총을 파괴하고 내성 병원균의 출현을 촉진할 수 있는 항생제 사용도 의료관련감염병 발생의 외부적 요인으로 들 수 있다. 셋째, 감염 시점부터 증상 발생까지 시간적인 지연이 있을 수 있다. 병원균에 노출된 직후에는 나타나지 않다가 입원기간 도중 또는 환자가 퇴원한 후에도 발생할 수 있다. 시기는 감염 유형과 환자의 상태(기저질환 유무)에 따라 다를 수 있으며, 퇴원 몇 주 후에 나타나기도 한다. 넷째, 심각한 합병증을 일으킬 수 있다. 장기 입원, 의료비 증가, 항생제내성률 및 사망률 증가를 포함한 심각한 합병증을 초래할 수 있다. 또한 항생제내성 병원균의 광범위한 전파로 다른 환자, 의료종사자나 방문객에게 위험을 초래할 수 있다. 다섯째, 적절한 감염 관리로 예방이 가능하다. 예방관리활동은 의료기관 내 항생제내성 전파 예방을 위한 의료기관 구성원에 대한 손 위생, 의료기관내 의료장비를 포함한 환경 청결 유지 및 소독 살균, 항생제의 적절한 사용, 필요한 경우 환자 격리 및 코호트 격리시행이 포함된다.

의료관련감염병은 항생제내성 병원균의 증가로 인하여 지역사회 내성균 확산으로 인한 질병 부담이 증가할 수 있어, 전세계적으로 중요한 문제점으로 대두되고 있다. 따라서, 국내 법정 의료관련감염병 예방을 위해서는 환례의 조기인지를 통한, 엄격한 감염 관리 지침을 적용, 지역사회 유행확산 등을 예방하는 것이 매우 중요하다.

4 국내 법정 의료관련감염병의 주요 특성

1. (전수감시) 반코마이신내성황색포도알균(VRSA) 감염증

2017년 6월부터 전수감시체계로 감시하고 있으며, 국내 VRSA 감염증은 현재까지 보고되지 않았다. 반코마이신중등도내성황색포도알균(VISA)은 2019년 5월부터 전수 감시 중이며, 전수감시 지정 이후 2024년 12월까지 총 19건이 신고되었다. 2021년 국가 항균제내성균 조사 연보에 따르면, 종합병원급 이상에서 분리된 황색포도알균의 약 45%에서 메티실린 내성이 확인되었으며, 반코마이신 항생제의 사용이 증가하면 향후 VRSA의 발생 가능성이 상존하는 상황이다. 감염증 환자 또는 병원체 보유자와의 직·간접접촉 및 오염된 의료기구, 환경 등을 통해서도 전파되는 것으로 알려져 있다. 균혈증, 피부 및 연조직 감염, 수술 부위 감염 등 다양한 감염증을 유발할 수 있으며, 감염부위는 여드름, 종기 또는 기타 피부질환처럼 나타날 수 있다.

2. (전수감시) 카바페넴내성장내세균목(CRE) 감염증

카바페넴계 항생제(ertapenem, meropenem, doripenem, imipenem) 중 최소 한 가지 이상에 내성인 장내세균목의 균종에 의한 감염질환이다. 2017년부터 전수감시체계로 전환하였으며, 2017년 5,717건에서 2020년 18,113건이 보고되고 있다. 대부분 단순 보균 상태이며, 이는 치료의 대상에 해당하지 않는다. 하지만, 감염증으로 인하여 증상이 나타나는 경우(요로감염, 위장관염, 폐렴, 패혈증 등)에는 항생제 치료의 대상이 되며, 다른 여러 계열 항생제에 내성을 나타내는 경우가 많아 치료에 어려움이 따른다.

3. (표본감시) 반코마이신내성장알균(VRE) 감염증

질병관리청 표본감시 결과에 따르면, 재원일수 1,000일당 VRE 혈액검체 분리율은 2012년 0.04건, 2022년 0.08건, 2024년 0.06건이었다. 장알균은 위장관과 비뇨생식계에 상재하고 정상인에서는 쉽게 감염을 일으키지 않지만, 노인, 면역저하 환자, 만성 기저질환자 또는 병원에 입원 중인 환자에서요로감염, 창상감염, 균혈증 등의 각종 기회감염증을 일으키며 감염부위에 따라 다양한 감염증을 유발한다.

4. (표본감시) 메티실린내성황색포도알균(MRSA) 감염증

질병관리청 표본감시 결과에 따르면, 재원일수 1,000일당 혈액검체 분리율은 2016년 0.13건에서 2020년 0.11건, 2024년 0.08건으로 꾸준히 감소하는 추세를 보였다. 면역력이 현저히 낮아진 경우 감염된 사람과 접촉하거나, 수건, 면도기와 같은 개인위생 용품을 함께 사용하는 경우 감염될 수 있다]. MRSA는 다양한 부위에 감염을 일으킬 수 있는데, 피부와 피하 조직 감염이 가장 흔하며, 골수염, 뇌수막염, 폐렴, 폐농양, 농흉과 같은 침습적 감염도 발생할 수 있다. 특히, MRSA로 인한 감염성 심내막염은 다른 병원균 감염에 비해 사망률이 높고, 정맥을 통한 약물 오남용과 밀접한 연관성이 있음이 알려져 있다.

5. (표본감시) 다제내성녹농균(MRPA) 감염증

카바페넴계, 아미노글리코사이드계, 플로로퀴놀론계 항생제에 모두 내성인 다제내성녹농균(MRPA)에 의한 감염 질환을 의미한다. 의료감염, 감염된 사람과의 직접 접촉이나 오염된 물체를 통한 전파가 주로 발생하며, 요로감염과 인공호흡기관련 폐렴 등 주요 의료관련 감염의 원인균이다. 감염부위에 따라 피부감염, 욕창, 폐렴, 균혈증, 패혈증, 수막염 등 다양한 감염증을 유발할 수 있다.

6. (표본감시) 다제내성아시네토박터바우마니균(MRAB) 감염증

카바페넴계, 아미노글리코사이드계, 플로로퀴놀론계 항생제에 모두 내성인 다제내성아시네토박터바우마니균(MRAB)에 의한 감염질환을 의미한다. 건강한 사람은 감염위험이 매우 적으나 면역저하자, 만성폐질환자, 당뇨환자는 감염에 취약하며 특히 인공호흡기구 사용환자, 장기간 입원환자는 감염의 위험성이 높다. 감염부위에 따라 폐렴, 혈류감염, 창상감염 등 다양한 감염증을 유발하는 것으로 알려져 있다.

참고 문헌

1. 질병관리본부, 의료관련감염관리학회. 의료관련감염 표준 관리지침. 2017.
2. 질병관리청. 국가 항균제내성균 조사연보. 2022.
3. 질병관리청. 법정감염병 진단검사 통합지침 4판. 2023.
4. 질병관리청. 의료관련감염병 관리지침. 2024.

5. 질병관리청. 전수감시 의료관련감염병 역사조사의 이해 교육자료. 2024.
6. 질병관리청. 표본신고감염병 감염병 의료관련감염병; 다제내성아시네토박터바우마니균(MRAB) 감염증 기간 별 통계. 질병관리청. [cited 2024 Dec 23]. Available form: https://dportal.kdca.go.kr/pot/is/st/hai.do.
7. 질병관리청[인터넷]. 의료감염관리 [Cited 2024 Dec 23]. Available form: https://www.kdca.go.kr/contents.es?mid=a20301080100.
8. 질병관리청[인터넷]. 전수감시감염병 감염병 제2급 반코마이신내성황색포다알균(VRSA) 감염병 기간별 통계 [Cited 2024 Dec 23]. Available form: https://dportal.kdca.go.kr/pot/is/summaryEDW.do.
9. 질병관리청[인터넷]. 전수감시감염병 감염병 제2급 카바페넴내성장내세균목(CRE) 감염병기간별 통계 [Cited 2024 Dec23]. Available form: https://dportal.kdca.go.kr/pot/is/summaryEDW.do.
10. 질병관리청[인터넷]. 표본신고감염병 감염병 의료관련감염병; 메티실린내성황색포도알균(MRSA) 감염증 기간별 통계. [Cited 2024 Dec 23]. Available form: https://dportal.kdca.go.kr/pot/is/st/hai.do.
11. Centers for Disease Control and Prevention (CDC) [Internet]. About Acinetobacter [Cited 2024 Dec 23]. Available form: https://www.cdc.gov/acinetobacter/about/?CDC_AAref_Val=https://www.cdc.gov/HAI/organisms/acinetobacter.html.
12. Centers for Disease Control and Prevention (CDC) [Internet]. About Pseudomonas aeruginosa. CDC. [Cited 2024 Dec 23]. Available form: https://www.cdc.gov/pseudomonas-aeruginosa/about/?CDC_AAref_Val=https://www.cdc.gov/hai/organisms/pseudomonas.html.
13. Centers for Disease Control and Prevention (CDC) [Internet]. Healthcare-Associated Infectons (HAIs) [Cited 2024 Dec 23]. Available form: CDC, https://www.cdc.gov/healthcare-associated-infections/php/data/index.html.
14. European Centre for Disease Prevention and Control (ECDC) [Internet]. Healthcare-associated infections. ECDC [Cited 2024 Dec 23]. Available form: https://www.ecdc.europa.eu/en/healthcare-associated-infections.
15. Garner JS, Jarvis WR, Emori TG, et al. CDC definitions for nosocomial infections, 1988. Am J Infect Control. 1988;16(3):128-40.
16. Richard P. Wenzel. Vancomycin-Resistant Gram-Positive Pathogens: Potential Approaches For Prevention And Control. In: Wenzel RP. Prevention and Control of Nosocomial Infections. 4th ed. 2003.
17. Siddiqui AH, Koirala J. Methicillin-Resistant Staphylococcus aureus. 2023. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025.
18. WHO. Report on the Burden of Endemic Health Care-Associated Infection Worldwide. 2012.

제2절 감염예방 및 관리의 종류와 적용범위

1 감염예방 및 관리의 정의 및 목적

감염예방 및 관리(infection prevention and control, IPC)는 환자 및 의료종사자의 감염을 보호하기 위한 조치로, 근거 기반 접근방식을 토대로 구성된 임상 및 공중 보건의 핵심분야이다. 감염예방 및 관리의 목적은 환자 및 의료 시설 방문자가 의료 서비스를 제공받거나 의료종사자가 의료서비스를 제공하는 과정에서 발생하는 항생제내성 병원균으로 인한 감염 및 발병을 통한 확산 등 예방할 수 있는 감염 위험을 차단함에 있다.

2 의료관련감염 및 의료관련감염병의 전파 방식

의료종사자, 환자, 의료 장비 또는 오염된 표면과의 직·간접적 접촉으로 인해 발생할 수 있다. 의료관련감염병의 전파 가능한 경로는 다음과 같다. 첫째, 직접접촉 전파(direct contact transmission)로 병원균은 신체 접촉(예: 의료종사자가 환자를 만지거나 환자 간 접촉)을 통해 감염되거나 병원균이 집락화된 사람에서 면역력이 취약한 사람에게 직접 전파될 수 있다. 예를 들어, 상처 드레싱, 체액(혈액, 대변) 또는 점막 표면과의 접촉은 메티실린내성황색포도알균(MRSA), 반코마이신내성장알균(VRE)과 같은 병원균을 전파시킬 수 있다. 둘째, 간접접촉 전파(indirect contact transmission)로, 오염된 물체 또는 의료 장비, 문 손잡이 또는 침대 린넨과 같은 표면을 통해 병원균이 전파되는 것을 포함한다. 예를 들어, 카테터, 인공호흡기, 온도계와 같은 오염된 표면 또는 의료기구는 녹농균(*P. aeruginosa*), 아시네토박터바우마니균(*Acinetobacter baumannii*) 등과 같은 다제내성균을 전파할 수 있다. 셋째, 비말 전파(droplet transmission)로, 기침이나 재채기에서 나오는 큰 호흡기 비말(respiratory droplet)이 다른 사람의 점막 표면(눈, 코 또는 입)에 직접 접촉하면서 전파된다. 넷째, 혈액 매개 전파(blood-

borne transmission)로 병원균이 포함된 혈액이나 기타 체액에 노출되어 전파되는 방식을 의미하며, 바늘 또는 날카로운 도구의 찔림 사고를 통해 발생한다.

3 의료관련감염 및 의료관련감염병 관리의 기본 원리

감염 사슬은 의료 환경에서 감염이 전파되는 과정을 이해하는 데 사용되는 모델이다. 감염이 한 사람에서 다른 사람으로 전파되는 방식을 설명하며, 감염 고리의 모든 사슬을 차단하는 것이 의료관련감염병의 확산을 예방하는 데 어떻게 도움이 될 수 있는지 알 수 있다. 감염 사슬은 다음과 같이 여섯 가지 핵심 요소로 구성된다(표 6-2-1, 그림 6-2-1).

표 6-2-1. 감염 사슬의 핵심 6요소

	정의 및 특징	예시	예방 및 관리
① 감염인자 (Microorganism)	질병을 유발하는 주 원인체로 병원체(Pathogen)로도 불린다.	박테리아, 바이러스, 진균 또는 기생충과 같은 감염을 일으키는 미생물	배양 또는 검사를 통한 병원체의 적절한 식별 및 진단, 적절한 항균 치료
② 병원소 (Reservoir/Source)	병원체가 생존하고, 번식하는 환경 또는 숙주를 의미한다.	전반적인 의료 환경, 장비, 체액(피, 타액), 소변/대변, 물/음식물, 피부, 호흡기	표면 세척 및 소독, 감염자 격리, 올바른 손씻기
③ 탈출구 (Port of Exit)	병원체가 병원소를 빠져나가는 경로이다.	피부 접촉, 피부 표면 접촉, 혈액, 체액, 구강 타액, 소변/대변, 관절액	마스크, 장갑, 가운과 같은 개인보호장비(PPE)를 사용, 적절한 상처 관리, 올바른 기침 에티켓
④ 전파방법 (Modes of Transport)	병원체가 병원소에서 감염 가능한 숙주로 전파되는 방식이다. 크게 직접전파와 간접전파로 분류된다.	직접전파: 공기(Airborne), 비말(Droplet), 접촉(Contact), 물림(bite), 찔림(Needlestick or other sharp injury) 간접전파: 비생체접촉매개물(fomites)	개인보호장비의 올바른 사용, 손 위생, 의료 장비 살균, 감염된 환자 격리 및 환경관리, 주의지침(표준주의, 접촉주의, 비말주의 등)의 엄격한 준수
⑤ 진입구 (Port of Entry)	병원체가 새로운 숙주에 들어가는 경로로, 병원체가 진입할 수 있는 신체의 진입구 모두를 포함한다.	신체 개구부(입, 눈, 비뇨기계, 호흡기계), 수술 상처, 정맥 카테터 등	시술 중 멸균 기술의 적절한 사용, 상처 봉합, 장갑과 마스크를 사용하여 병원체 진입을 차단

	정의 및 특징	예시	예방 및 관리
⑥ 감수성 있는 숙주 (Susceptible host)	면역 체계 약화, 기저질환, 나이 등의 요인, 또는 의료종사자 등 감염될 위험이 있는 사람이다.	노인 환자, 신생아, 만성 질환자 (예: 당뇨병, 암), 면역억제 요법을 시행 중인 환자, 의료종사자	예방 접종을 통해 숙주의 면역력 강화, 환자의 영양상태 최적화, 예방적 항생제나 항진균제의 사용

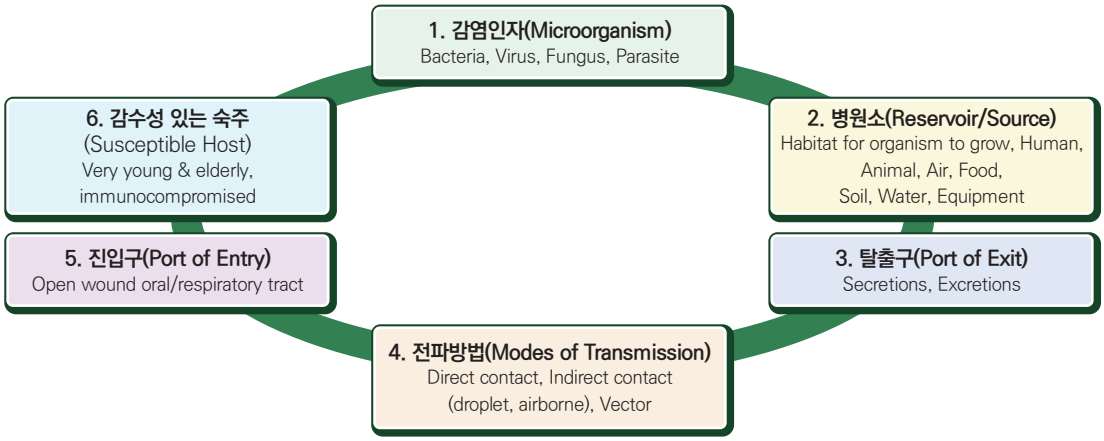


그림 6-2-1. 감염예방 및 관리를 위한 주의지침 및 종류

4 감염예방 및 관리를 위한 주의지침 및 종류

1. 표준주의

표준주의(Standard precautions)란 환자, 보호자, 의료기관 방문객과 의료종사자의 감염으로부터의 안전을 보장하기 위하여, 감염 여부와 관계없이 모든 환자에게, 모든 의료기관에서, 모든 의료종사자가 항상 적용해야 하는 지침이다. 표준주의를 통하여 감염원(혈액 및 기타 체액, 분비물 또는 배설물, 손상된 피부나 점막 등 오염될 가능성이 있는 의료 환경 내의 장비나 물품)으로부터 감염 연결고리(그림 6-2-1)에 개입함에 따라 전파 위험을 낮출 수 있다. 국내 의료관련감염 표준 예방지침의 권고 사항으로는 병동과 외래의 대기 장소에는 손위생과 관련된 물품을 제공하고, 손위생 방법을 안내하도록 하는 등 구체적인 지침을 제시하고 있다. 또한, 환자와 가까운 거리에 있거나, 환자가 만지는 물건 및 환경 표면은 자주 청소 및 소독하고, 다른 사람에게 감염을 전파할 위험이 있는 환자의 경우 전파 가능성

을 고려하여 가능한 1인실에 격리하도록 권고하고 있다.

2. 전파경로별 주의지침

표준주의만으로는 특정 병원체의 전파를 예방하기에 충분하지 않을 수 있으며, 특정 병원체에 대한 감염 또는 집락화가 확인 또는 의심되는 환자를 치료 및 간호함에 있어 전파경로별 주의지침이 필요하다. 전파경로별 주의지침(Transmission based precautions)은 병원체의 전파경로를 기준으로 구분하며, 접촉주의, 비말주의, 공기주의로 구분할 수 있다.

먼저, 접촉주의(contact precautions)는 환자 또는 주변 환경과의 직접 또는 간접적인 접촉(환자 치료에 사용되는 모든 의료장비)으로 병원체가 전파되는 것을 예방하고 관리하기 위해 사용된다. 광범위한 상처 배농, 변실금, 또는 기타 신체 분비물이 있는 병원체의 광범위한 오염 및 전파 위험이 높은 경우에도 접촉 주의를 통하여 조치를 취한다. 국내 의료관련관리 지침에서는 접촉주의가 필요시 가능한 1인실로 입원해야 하며, 감염병 전파 가능성이 높은 환자를 우선 배치한다. 1인실이 여유가 없는 경우, 동일한 병원체에 감염되었거나, 보균 중인 환자들끼리는 한 병실에 입원(코호트)할 수 있음을 권고하고 있다. 또한, 접촉주의가 필요한 환자를 직접 접촉하거나, 환자 주변의 물건을 만져야 할 때에는 손위생 수행 후 장갑을 착용하고, 옷이 오염될 것으로 예상될 때에는 가운을 착용한다. 접촉주의에 필요한 개인보호구는 병실 입구에서 제공되어야 한다. 병실을 나올 때에는 장갑과 가운을 벗어 의료폐기물통에 버리고 손위생을 수행하도록 권고하고 있다. 치료장비와 기구관리로서, 환자의 치료를 위해 필요한 물품은 가능한 일회용품을 사용하고, 다른 환자와 공유해서 사용하지 않도록 하며, 접촉주의는 환자 퇴실 후 병실 청소를 완료할 때까지 유지하도록 권고하고 있다.

비말주의(droplet precautions)는 감염된 환자 호흡기 분비물과의 밀접한 호흡기 또는 점막 접촉을 통하여 병원균이 전파되는 것을 예방하고 관리하기 위해 조치한다. 비말주의가 필요한 환자는 가능한 1인실에 배치한다. 다만, 비말주의가 필요한 병원균은 의료시설 내 장거리 전파가 불가능하기 때문에, 이를 예방하기 위한 음압병실 등의 특수 공기 처리 및 환기 시스템이 있는 병실 사용은 필요하지 않다.

국내 지침에서는 환자 배치에 있어, 비말주의가 필요한 환자는 가능한 1인실에 배치하고, 1인실 사용이 제한이 있는 경우, 일반 병실에서 코호트 구성 시 동일한 병원체에 감염된 환자들로 배치하도록 권장한다. 코호트 격리 시 병상 간 이격거리는 1 m^3 이상 유지하고, 접촉 기회를 줄이기 위해 침대 사이에 물리적 칸막이를 설치하도록 하고 있다. 비말주의가 필요한 환자의 병실에 들어갈 때에는 수술용 마스크를 개인보호구로서 착용한다.

공기주의 지침은 공기 중 부유하면서 장기적으로 생존할 수 있는 병원체로부터 예방하고 관리하기

위해 사용한다. 공기주의가 필요한 환자는 감염 격리실(airborne infection isolation room)에 격리한다. 가능한 해당 병원체에 대한 예방접종력이 없거나, 면역력이 저하된 의료진은 치료 및 간호에 참여하지 않고, 감염 환자를 접촉하지 않도록 한다. 국내 표준예방지침에서는, 공기주의가 필요한 환자는 음압격리실에 배치하며, 음압격리실은 환자 개별 사용하는 화장실, 세면대, 샤워실이 있어야 하며, 의료진의 손위생을 위한 시설이 갖추어져 있어야 한다. 외래에서는 공기전파가 가능한 환자를 선별하기 위한 체계를 구축하도록 권고하고 있다. 음압격리실의 시설기준으로는, 최소 6회 이상의 공기가 순환되도록 하며, 신규설비의 경우 12회 이상을 권장한다. 공기는 곧바로 건물 밖으로 배출되도록 하거나 헤파필터가 있는 공조시스템을 통과해야 한다. 출입 시 외에는 문은 항상 닫혀있어야 한다. 음압격리실 방의 한 측면에서 공기가 들어올 시 환자의 침대를 지나 방의 반대쪽으로 공기가 흐르도록 해야 한다. 음압격리실의 방과 외부의 기압은 최소 2.5 Pa 이상 차이나도록 한다. 매일 육안으로 관찰할 수 있는 지표로 공기의 압력 상태를 확인한다. 외부의 공기가 들어오지 않도록 방은 잘 밀폐되어 있도록 한다. 공기로 전파되는 병원체 감염이 의심되거나 확진 환자의 치료영역으로 들어갈 시에는 N95 마스크를 착용하여 제대로 착용이 되었는지 확인한다. 백신으로 예방이 가능한 공기전파 감염병 환자를 치료 및 간호 시 면역 형성이 없는 의료종사자는 업무배제가 되도록 권장하고 있다.

다음은 각 전파경로별 주의지침이 적용되는 감염 병원체의 종류이다. 일부 병원체는 두 가지 이상의 경로를 통해 전파될 수 있으므로, 병원체의 전파경로를 면밀하게 파악하는 것이 중요하다.

표 6-2-2. 의료관련감염 및 의료관련감염병의 전파경로별 주의지침과 관련 병원체의 종류

전파경로별 주의지침	감염 병원체 종류
접촉주의 (Contact precautions)	Adenovirus, Amebiasis, Anthrax, Cholera, Clostridium difficile, Congenital rubella, Cutaneous Diphtheria, Rotavirus, Herpes simplex virus (Mucocutaneous, disseminated or primary), Herpes zoster virus (Disseminated disease in any patient), Impetigo, Lice (Head, Pediculosis), **Multidrug-resistant organisms (infection or colonization)**, Parainfluenza virus (in infants and young children), Respiratory syncytial virus (in infants and young children), S.aureus (Staphylococcal scaled skin syndrome), Scabies, Severe acute respiratory syndrome (SARS), Smallpox, Group A streptococcus, M. Tuberculosis (Extrapulmonary, draining lesion), Varicella Zoster virus, Lassa fever virus, Marburg fever virus, Crimen-congo fever virus

전파경로별 주의지침	감염 병원체 종류
비말주의 (Droplet precautions)	Adenovirus, Anthrax, Avian influenza, Diphtheria (Pharyngeal), Haemophilus influenza type b (Epiglottitis), Meningococcal disease (Sepsis, Pneumoniae, Meningitis), Mumps, Mycoplasma pneumoniae, Parvovirus B19, Pertussis, Plaque (Pneumonic), Group A streptococcus, Rhinovirus, Rubella, Severe acute respiratory syndrome (SARS), Lassa fever virus, Marburg fever virus, Crimen-congo fever virus
공기주의 (Airborne precautions)	Herpes zoster virus (Disseminated disease in any patient), Meales, M.tuberculosis, Severe acute respiratory syndrome (SARS), Smallpox, Varicella Zoster

참고 문헌

1. 질병관리본부, 대한의료관련감염관리학회. 의료관련감염 표준 관리지침, 2017.
2. Centers for Disease Control and Prevention (CDC) [Internet]. Bloodborne infectious Disease Risk Factors. CDC [Cited 2024 Dec 23]. Available form: https://www.cdc.gov/niosh/healthcare/risk-factors/bloodborne-infectious-diseases.html.
3. Centers for Disease Control and Prevention (CDC). Guideline for Isolation Precautions: Preventing Transmission of Infectious Agents in Healthcare Settings. 2007.
4. Fennelly KP. Particle sizes of infectious aerosols: implications for infection control. Lancet Respir Med. 2020;8(9):914-24.
5. Hota B. Contamination, disinfection, and cross-colonization: are hospital surfaces reservoirs for nosocomial infection? Clin Infect Dis. 2004;39(8):1182-9.
6. National Health Service (NHS) [Internet]. Chapter 1: Standard infection control precautions (SICPs). NHS [Cited 2024 Dec 23]. Available form: https://www.england. nhs.uk/national-infection-prevention-and-control-manual-nipcm-for-england/chapter-1-standard-infec-tion-control-precautions-sicps/.
7. National Health Service (NHS) [Internet]. Transmission-Based Precautions, NHS [Cited 2024 Dec 23]. Availible form: https://www.cdc.gov/infection-control/hcp/basics/transmission-based-precautions.html.
8. Pittet D, Hugonnet S, Harbarth S, et al. Effectiveness of a hospital-wide programme to improve compliance with hand hygiene. Infection Control Programme. Lancet. 2000;356(9238):1307-12.
9. WHO. Global report on infection prevention and control. 2024.

제3절 의료관련감염병 감시체계

감시활동이란 감시대상인 해당 건강 관련 사건에 관한 자료를 지속적이고 체계적으로 수집, 분석, 해석 및 배포하여 감시활동이 수행 중인 지역 내 질병이환 및 사망률을 줄이고 건강을 개선하는 것이다.

1 의료관련 감염병 감시 대상

1. 감시대상

2011년 1월부터 44개 상급종합병원을 대상으로 의료관련감염병(다제내성균 6종: VRSA, VRE, MRSA, MRPA, MRAB, CRE) 표본감시체계 운영하였으며, 표본 감시기관을 2011년 7월 100개소, 2024년 363개소로 확대 지정하였다. 2012년 9월에는 의료관련감염병 관련 고시 개정을 통한 환자와 병원체 보유자 정의 개정 및 산출지표 제정이 이루어졌으며, VRE와 MRSA의 신고대상 검체를 혈액에서 혈액 외 검체까지 확대하였다. 표본감시를 운영하던 CRE, VRSA는 2017년 6월부터 전수감시체계로 전환하였다.

2 의료관련감염병 감시체계 및 발생현황

1. 전수감시 대상 2종

의료기관에서는 2종의 의료관련감염병(VRSA, CRE) 발생 시 방역통합정보시스템을 통하여 보건소로

신고하면, 관할 보건소는 시스템을 이용하여, 시·도 보건담당 부서로 보고하고, 시·도에서는 관련정보를 확인하여 질병관리청으로 보고하도록 한다. 환자와 병원체보유자로 구분하여 신고, 집계되며 환자의 임상증상 유무와 상관없이 검체 종류에 따라 환자, 병원체보유자를 분류한다. 신고를 위한 각 진단기준은 다음과 같다.

VRSA(VISA), CRE 신고를 위한 진단기준

환자란 혈액에서 해당 다제내성균이 분리된 자로 정의하며, 병원체 보유자란 혈액 이외 임상검체에서 해당 다제내성균이 분리된 자로 정의한다. 여기서 임상검체는 객담, 소변, 대변(직장도말 포함), 피부, 상처, 농양, 뇌척수액, 기관흡인액, 체액(흉막액/복막액/심낭액) 등을 포함한다.

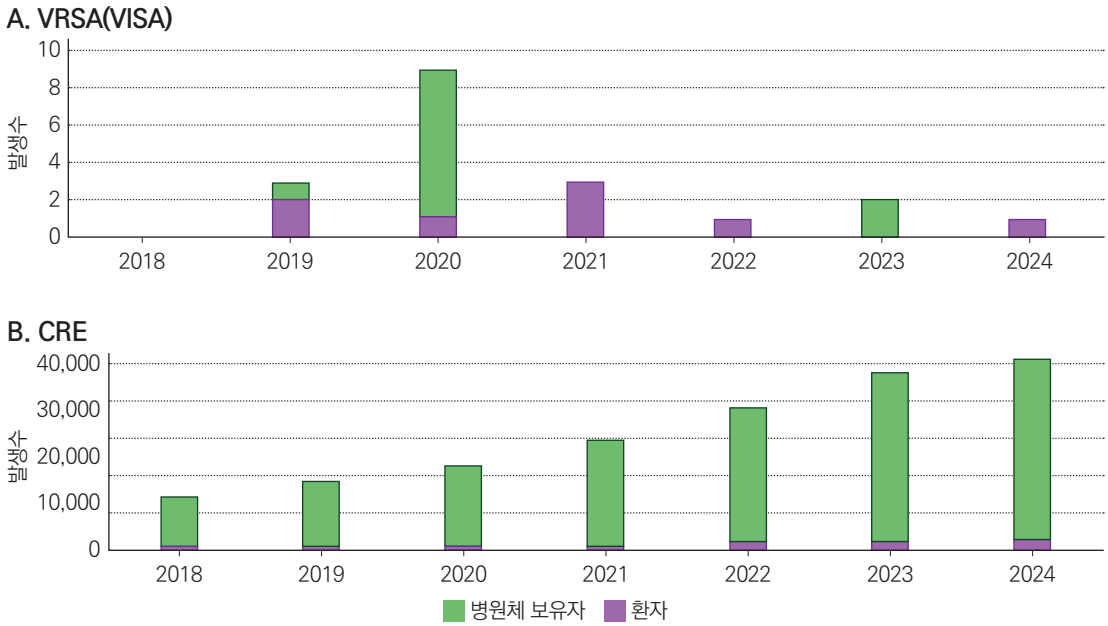


그림 6-3-1. 국내 전수감시 대상 2종 환자 및 병원체보유자 발생 현황(VRSA, CRE, 2024년 비확정통계)
반코마이신내성황색포도알균(VRSA) 감염증은 「감염병의 진단기준」 고시 개정(2019.4.30.)에 따라 반코마이신중등도 내성황색포도알균(VISA) 감염증을 포함하며, 모두 VISA 감염증으로 신고된 자료임

2. 표본감시 대상 4종

표본감시대상으로 지정된 의료기관에서는 4종의 의료관련감염병(VRE, MRSA, MRPA, MRAB) 발생

시 방역통합정보시스템을 통하여 보건소로 신고하면, 관할 보건소는 시스템을 이용하여, 시·도 보건담당 부서로 보고하고, 시·도에서는 관련정보를 확인하여 질병관리청으로 보고하도록 한다(그림 6-3-2).

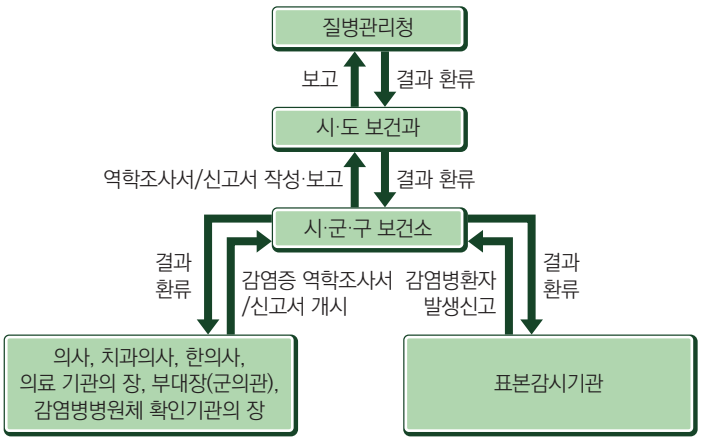


그림 6-3-2. 표본감시체계

표본감시대상 4종 산출방법(VRE, MRSA, MRPA, MRAB)

① 표본감시 의료기관은 감염관리실 설치 의료기관 중 종합병원(상급종합병원 포함), 100병상 이상 병원, 공공병원을 기준으로 지정한다.

② 혈액 검체 분리율과 혈액외 검체 분리율로 구분하여 매주 신고, 집계되며, 각 산출식은 다음과 같다.

$$\text{혈액 검체 분리율} = \frac{\text{혈액 검체에서 해당 다제내성균이 분리된 자}}{\text{총 재원일수}} \times 1{,}000$$

$$\text{혈액외 검체 분리율} = \frac{\text{혈액외 검체*에서 해당 다제내성균이 분리된 자}}{\text{총 재원일수}} \times 1{,}000$$

* 임상검체는 객담, 소변, 대변(직장도말 포함), 피부, 상처, 농양, 뇌척수액, 기관흡인액, 체액(흉막액/복막액/심낭액) 등을 포함한다.

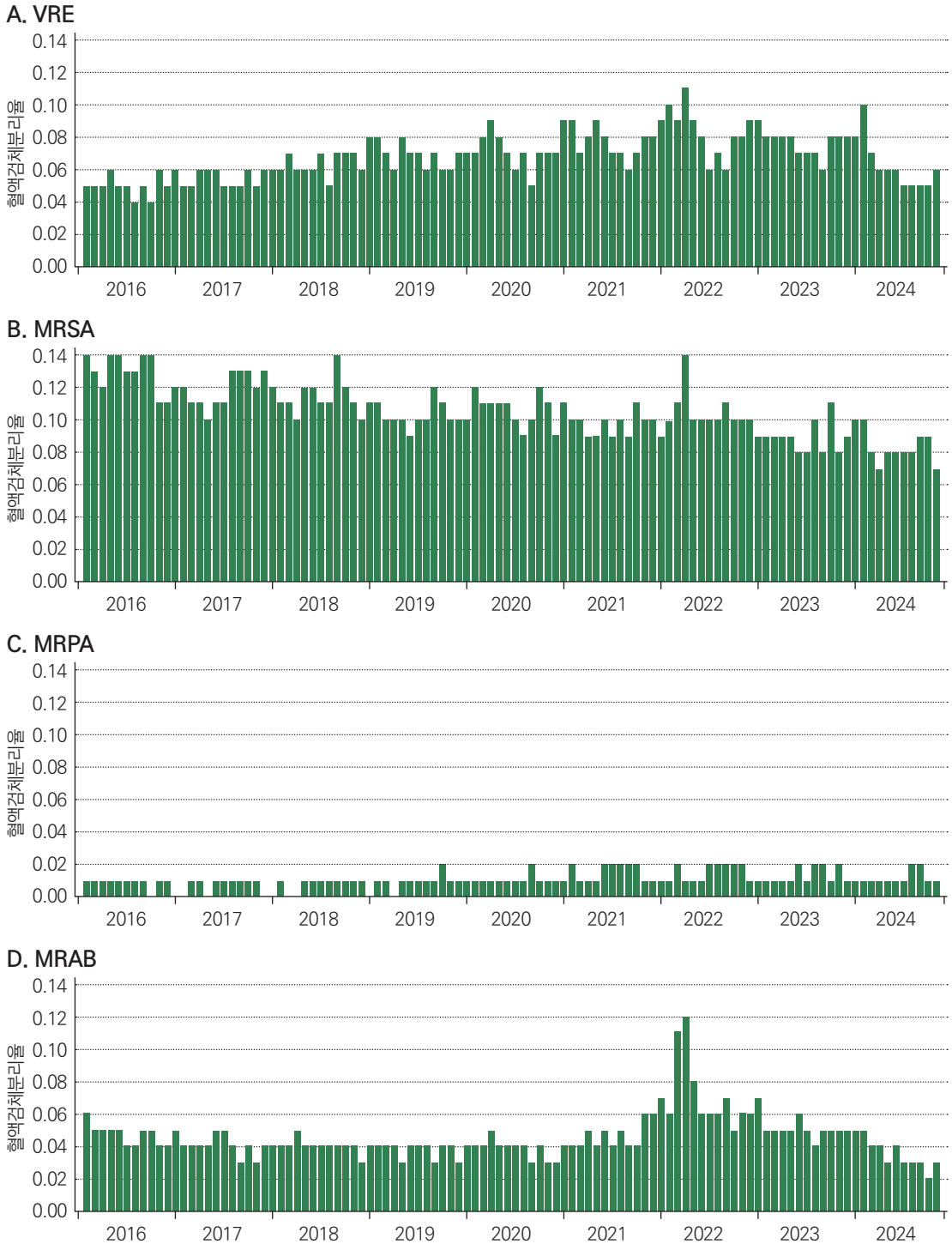


그림 6-3-3. 국내 표본감시 대상 4종 혈액검체분리율 현황(VRE, MRSA, MRPA, MRAB)

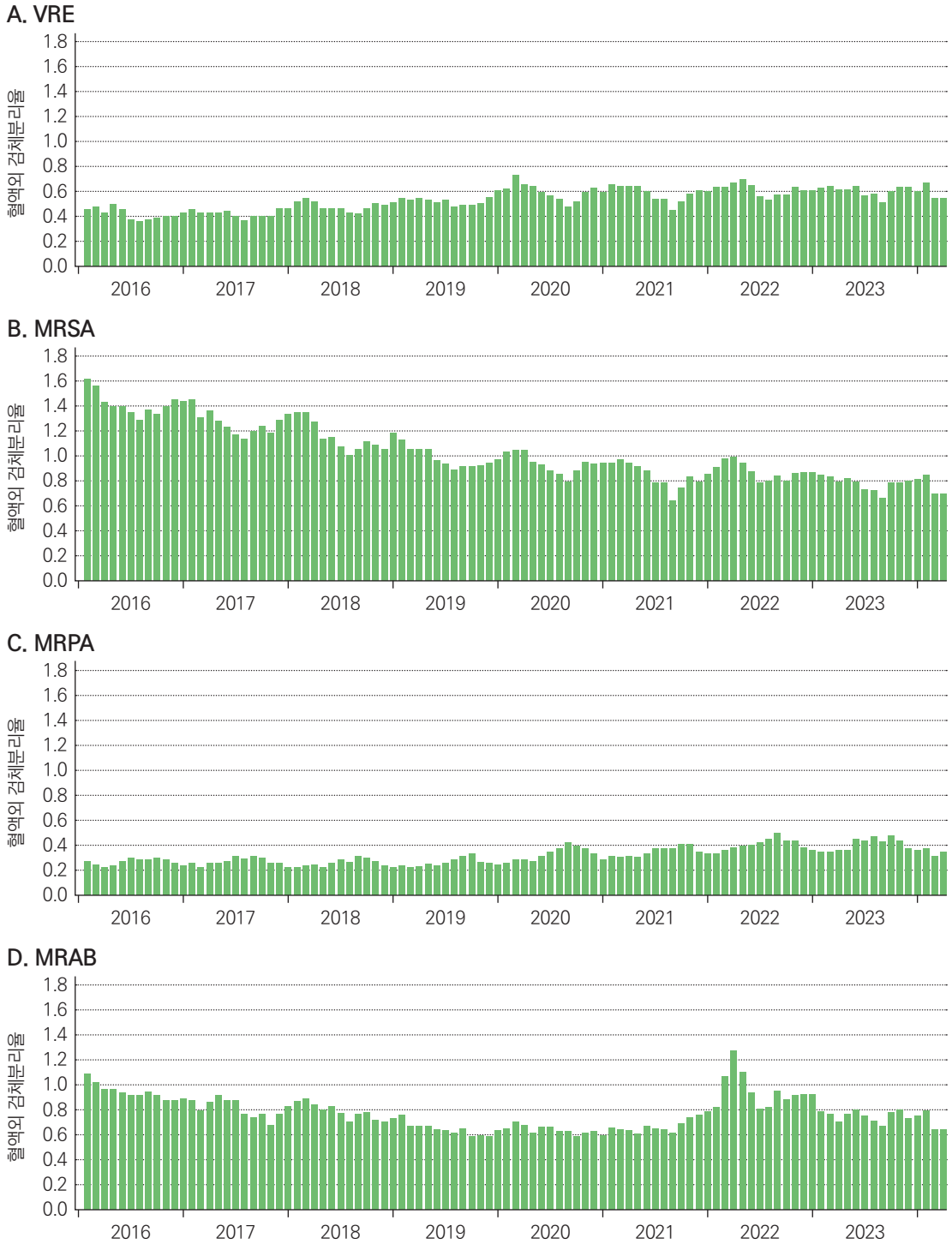


그림 6-3-4. 국내 표본감시 대상 4종 혈액외 검체분리율 현황(VRE, MRSA, MRPA, MRAB)

참고 문헌

1. 질병관리청. 의료관련감염병 관리지침. 2024.
2. 질병관리청[인터넷]. 의료감염관리 [Cited 2024 Dec 23]. Available from: https://www.kdca.go.kr/contents.es?mid=a20301080100.
3. 질병관리청[인터넷]. 표본신고감염병 의료관련감염병 기간별 통계 [Cited 2024 Dec 23]. Available from: https://dportal.kdca.go.kr/pot/is/st/hai.do.
4. 질병관리청[인터넷]. 전수감시감염병 제2급 주요통계(기간별) [Cited 2024 Dec 23]. Available from: https://dportal.kdca.go.kr/pot/is/summaryEDW.do.
5. German RR, Lee LM, Horan JM, et al. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Recomm Rep. 2001;50(RR-13):1-35; quiz CE1-7.

제4절 의료관련감염 및 의료관련감염병 유행 시 역학조사 단계와 방법

1 의료관련감염 및 의료관련감염병 역학조사의 기본 개념

1. 서론

의료관련감염병(항생제내성 6종)은 감염병 예방법 제2조 제3호 및 제12호에 따라, 현재 전수(2종), 표본(4종) 감시대상 감염병이다. 의료관련감염병은 의료관련감염(의료법 제 4조에 따른 의료관련감염의 정의)이 주전파 경로이기 때문에, 의료관련감염병 역학조사 및 관리에 있어 의료관련감염과 따로 분리하여 적용하기에는 어려움이 존재한다. 본 장에서는 주로 의료관련감염의 주요 전파 경로와 역학조사 접근등에 대하여 살펴보고 고찰해 보고자 한다. 상세한 의료관련감염병 역학조사와 관련된 내용은 의료관련감염병 관리지침(질병관리청 2025)을 참고한다.

의료감염은 의료행위와 관련된 모든 감염을 의미하며, 의료환경이라는 용어는 의료 서비스가 제공되는 다양한 장소를 포괄적으로 지칭한다. 여기에는 급성기 병원, 응급의료센터, 혈액투석실, 치과, 외래주사실, 내시경실, 통증클리닉, 외래수술센터와 같은 특수한 장소도 포함된다. 이처럼 다양한 의료환경에서는 특정 장소나 서비스가 역학조사의 중심이 될 수 있다. 예를 들어, 급성기 병원은 환자 분류 및 응급 치료, 입원 및 외래 수술 절차, 면역 억제 환자의 관리, 재활 서비스, 중환자실과 같은 다양한 전문 영역을 포함하는 매우 복잡한 조직이다. 이와 같이 특정 환경에서 제공되는 환자의 유형(예: 암 환자, 장기 이식 환자, 투석 환자)과 임상 서비스(예: 내시경, 투석, 정맥주사)에 대한 이해는 감염병 전파 위험을 평가하고 예방 대책을 수립하는 데 중요하다.

한편, 의료환경에서 발생하는 문제는 감염병 유행뿐만 아니라 제조 과정에서 오염된 약물이나 기기의 사용과 관련될 수도 있다.

이와 같은 다양한 요인들은 의료환경에서의 감염관리와 역학조사의 중요성을 강조하며, 이를 통해 의료 환경 내 환자와 의료진의 안전을 보장할 수 있다.

2. 의료관련감염 및 의료관련감염병의 주요 원인

의료관련감염의 주요 원인은 병원체, 환경 및 의료행위, 그리고 환자 관련 요인으로 나눌 수 있다.

첫 번째로 병원체 요인은 주로 세균, 진균, 그리고 바이러스로 구분된다. 세균 중에서는 황색포도알균(*Staphylococcus aureus*), 특히 메티실린 내성 황색포도알균(MRSA), 대장균(*E. coli*), 녹농균(*Pseudomonas aeruginosa*), 아시네토박터(Acinetobacter) 종, 장알균(Enterococcus) 종이 주요 원인균으로 꼽힌다.

두 번째로 환경 및 의료행위 요인에는 침습적 의료기기의 사용, 수술 및 기타 침습적 시술, 부적절한 손 위생, 오염된 의료기기나 환경 표면, 그리고 항생제 과다 사용으로 인한 내성균 발생이 포함된다. 중심정맥관, 요로카테터, 인공호흡기 등 침습적 의료기기의 사용은 특히 의료관련감염 발생 위험을 높이는 요인으로 알려져 있다.

세 번째로 환자와 관련된 요인에는 면역력 저하, 장기 입원, 그리고 기저질환의 존재가 있습니다. 노인, 만성질환자, 면역억제제를 사용하는 환자는 감염 위험이 더 높다.

의료관련감염 및 의료관련감염병은 이러한 요인들이 복합적으로 작용하여 발생한다. 특히 낮은 면역력, 침습적 시술, 의료인의 실수, 그리고 입원 환경이 감염 발생의 주요 요인으로 작용하며, 장기 입원 환자나 침습적 의료기기를 사용하는 환자에게 발생 위험이 더욱 높다.

3. 의료관련감염 및 의료관련감염병의 주요 경로

환자들은 의료 서비스를 받는 과정에서 다양한 감염성 병원체에 노출될 위험이 있으며, 이러한 병원체는 여러 경로를 통해 환자에게 전파될 수 있다.

첫 번째로, 의료인이나 다른 환자로부터의 교차 감염이 주요 원인 중 하나다. 이는 의료진이 적절한 손 위생이나 개인보호구를 착용하지 않을 경우 발생할 가능성이 높다.

두 번째로, 환자가 이미 보유하고 있는 미생물이 원인이 되는 내인성 감염이 있다. 이는 환자의 면역상태가 약화되거나 의료적 개입으로 인해 내부 미생물이 병원성이 증가할 때 발생할 수 있다.

세 번째로, 병원 환경 표면이나 물체의 오염이 감염의 경로가 될 수 있다. 침대 난간, 욕실 표면, 배수구와 같은 장소에서 발견되는 미생물이 이에 해당한다.

네 번째로, 중심정맥관, 요로 카테터, 내시경, 수술 도구, 인공호흡기 등 오염된 의료기기를 통해 병원체가 전파될 수 있다. 이러한 기기들은 정기적인 멸균 및 관리가 필요하다.

다섯 번째로, 오염된 약물도 감염의 경로가 될 수 있다. 약물 제조 과정이나 보관 중에 병원체가 혼입될 경우 환자에게 심각한 영향을 미칠 수 있다.

이러한 다양한 경로는 의료 환경에서 감염 발생의 주요 요인이 되며, 이를 예방하기 위해 적절한 감염관리가 중요하다.

4. 의료기관에서 역학조사를 수행할 때 조사할 주요 자료원

의료기관에서 역학조사를 수행할 때는 감염병의 원인, 전파 경로, 그리고 영향받은 인구를 파악하여 효과적인 대책을 마련하기 위해 다양한 자료를 체계적으로 수집해야 한다. 주요 자료원은 다음과 같다.

먼저 환자와 관련된 정보로는 환자의 진단서, 병력 기록, 입원 및 퇴원 요약, 영상기록 등이 포함된다. 또한 증상 발현 날짜 및 시간, 신체 검사 결과, 치료 및 약물 투여 내역과 같은 임상 데이터도 중요하다. 실험실 검사 결과로는 혈액, 소변, 조직 등의 배양 검사 결과, 항생제 감수성 검사, 그리고 유전자형 검사 결과가 포함된다. 환자의 노출 정보로는 수술 또는 시술 이력, 침습적 의료기기 사용 여부와 기간, 투약 기록 등이 포함된다.

다음으로는 의료 환경 및 관리와 관련된 정보이다. 병실 배치 정보는 환자 이동 기록과 병실 및 병동 내 위치 정보(가능하다면 평면도 및 공조시스템 도면 포함)가 포함된다. 또한 공유 의료기기 사용 내역, 멸균 및 소독 기록과 같은 의료기기 및 장비 사용 기록도 필수적이다. 병원 표면, 물 및 공기질 검사 결과와 같은 환경 검체 데이터와 병원의 청소 및 소독 일정, 사용된 소독제와 농도 등의 환경 관리 기록도 포함된다.

의료진과 병원 운영에 관한 정보도 중요한 자료원이다. 의료진의 근무 기록, 건강 상태 및 감염 이력, 손 위생과 개인보호구(PPE) 착용 준수 기록이 이에 포함된다. 병원의 감염관리 정책 및 프로토콜, 감염관리 교육 기록, 환자와 의료진 간 접촉 빈도와 방식 등 병원 운영 데이터도 포함된다.

감염관리와 감시에 관한 데이터로는 병원 내 감시 데이터, 외부 보고 자료, 그리고 유행곡선 및 사례 라인 리스트 등이 있다.

마지막으로 기타 자료원으로는 약물 및 수액 관련 데이터와 환자 인터뷰가 있다. 약물이나 수액의 제조 및 유통 경로, 오염 여부 확인과 관련된 데이터가 중요한 역할을 한다. 환자 인터뷰를 통해 입원 전 활동 및 동선, 감염 가능성이 있는 접촉 또는 노출 이력을 확인할 수 있다.

이러한 다양한 자료원들은 의료기관 내 감염병 유행 시 효과적인 역학조사를 가능하게 한다.

2 의료관련감염 및 의료관련감염병 유행의 인지와 보고

의료관련감염 및 의료관련감염병 유행의 인지 과정은 감염병 유행 초기 단계에서 신속하고 정확한 판단을 내리는 것이 중요하며, 이는 감염 확산을 최소화하고 환자 및 의료진의 안전을 보장하는 데 핵심적인 역할을 한다. 이러한 과정은 비정상적인 감염 발생을 조기에 감지하고, 적시에 통제 조치를 시행함으로써 병원의 정상적인 운영을 유지하고 공중보건 위기를 예방할 수 있도록 돕는다.

더 나아가, 체계적인 데이터 수집과 분석을 통해 감염병 유행의 정확한 원인을 규명할 수 있으며, 이를 기반으로 유사한 상황이 반복되지 않도록 예방 대책을 마련하는 데 기여한다. 이와 같은 과정을 통해 의료기관은 감염병 유행 상황에서도 신속히 대응할 수 있는 역량을 갖추게 되며, 환자 치료의 연속성을 유지하고 공중보건의 안전성을 강화할 수 있다.

1. 유행 인지의 주요 방법

1) 감시 시스템(기존 의료관련감염병 및 의료관련감염 발생에 대해 감시가 진행되고 있었음을 전제)

(1) 일상적인 감염감시를 통해 비정상적 증가 탐지

비정상적 증가를 탐지하기 위한 방법은 통계적 기법, 임상 및 실험실 관찰, 그리고 자동화된 경보 시스템의 세 가지 주요 접근법으로 구분할 수 있다.

첫 번째로, 통계적 기법은 비정상적인 감염 발생 패턴을 파악하는 데 효과적이다. 관리도(control chart)와 유행곡선(epidemic curve) 같은 유행 탐지 알고리즘은 감염 발생의 비정상적 증가를 식별하는 데 활용된다. 또한 과거 데이터와 현재 발생률을 비교하는 비교 분석 방법도 유행을 탐지하는 데 중요한 역할을 한다. 이를 통해 시간별, 병동별, 환자군별 데이터를 분석하고, 특정 지역이나 집단에서 발생하는 클러스터(clusters)를 발견할 수 있다.

두 번째로, 임상 및 실험실 관찰은 의료진과 실험실 데이터의 관점을 결합하여 유행 가능성을 확인한다. 의료진이 동일한 병동이나 치료 과정에서 유사한 감염 사례를 반복적으로 발견할 경우 이는 유행의 신호로 간주될 수 있다. 또한 실험실 검사 결과에서 동일한 병원체가 여러 샘플에서 반복적으로 검출된다면, 이는 감염 확산의 가능성을 시사한다.

세 번째로, 자동화된 경보 시스템은 실시간 데이터를 기반으로 신속하게 비정상적인 감염 증가를 탐지한다. 병원 내 전자 감시 시스템은 감염 발생 데이터를 실시간으로 분석하여 위험 수준을 초과하는 패턴을 감지하면 자동으로 경보를 발송한다. 이러한 시스템은 감염률의 변화, 항생제내성 패턴의 변화,

특정 병원체의 보고 증가 등을 모니터링하여 조기 대응을 가능하게 한다.

이러한 다양한 방법들은 병원 환경에서 감염의 비정상적 증가를 조기에 탐지하고, 신속하게 대응하여 감염 확산을 최소화하는 데 기여한다.

(2) 감염률, 항생제내성 패턴 변화, 특정 병원체의 보고 증가 등

첫 번째로, 의료기관이 통상적인 감염률이나 항생제내성 패턴에 대한 기존 자료를 보유하고 있는 경우, 이를 현재 발생 상황과 비교함으로써 비정상적인 변화를 감지할 수 있다. 예를 들어, 특정 병원체의 발생률이 과거 데이터에 비해 유의미하게 증가하거나, 항생제내성률이 예상 범위를 초과하면 유행 가능성을 시사할 수 있다.

두 번째로, 기존 자료가 없는 경우에는 후향적으로 데이터를 수집하여 비교 분석을 진행한다. 이를 위해 과거 일정 기간 동안의 감염 사례를 검토하고, 유사한 상황에서의 데이터 패턴을 분석하여 현재의 발생률과 비교한다. 이러한 접근법은 새로운 감염 패턴이나 병원체의 발생 증가를 탐지하는 데 유용하다.

2) 보고 시스템

감염관리 담당자, 의료진, 실험실에서 감지한 구체적인 사례 확인을 통해 강화된 감염관리 조치를 시행함으로써, 추가적 감염 확산 예방이 가능하다.

(1) 감염관리 담당자가 감지한 사례

- (의료관련감염병) 사례: 중환자실에서 다제내성균 확산 발견
- 상황: 감염관리실에서 중환자실 데이터를 분석하던 중, 특정 병원균(예: CRE, Carbapenem-Resistant Enterobacterales)이 지속적으로 검출되는 것을 확인하였다. 이 병원균의 감염률이 통상적인 발생률(baseline)을 초과한 것으로 나타났다.
- 보고 절차: 감염 사례와 증가 추이에 대해 병원 관리자와 중환자실 담당 의료진에게 보고한다. 의료기관의 장은 관할 보건소에 감염병 발생신고를 해야하며, 보건소는 개별사례 역학조사를 시행한다. 또한, 시도는 보건소 보고내역을 통해 집단 역학조사 대상 여부를 판단한다.

(2) 의료진이 감지한 사례

- (의료관련감염) 사례: 수술 후 감염 사례
- 상황: 외과의사가 환자의 수술 후 상처 부위에서 비정상적인 염증과 삼출액을 관찰하였고, 같은 병동에서 유사 증상을 보이는 환자 2명이 추가로 확인되었다.

- 보고 절차: 해당 사례를 감염관리실에 보고하여 사례정의(case definition)를 검토하고 추가 조사를 요청한다. 감염원으로 의심되는 수술 장비나 환경 오염 여부를 확인하기 위해 역학조사팀이 소집된다.

(3) 실험실에서 감지한 사례

- (의료관련감염) 사례: 혈액 배양 검사에서 희귀 병원체 검출
- 상황: 병원 실험실에서 혈액 배양 검사 중 다수의 샘플에서 Candida auris와 같은 항생제내성이 높은 병원체가 반복적으로 검출되었다. 동일 병동에 입원 중인 환자들 사이에서 병원체 전파 가능성이 의심되었다.
- 보고 절차: 실험실 검사자가 즉시 감염관리실에 해당 결과를 보고한다. 이후 감염관리실은 관련 데이터를 분석하고, 감염 경로 및 확산 방지를 위한 조치를 시행한다.

이와 같은 체계적 보고와 대응은 병원 내 감염 발생을 조기에 차단하고, 환자 및 의료진을 감염으로부터 보호하며, 추가적인 유행을 예방하는 데 중요한 기초가 된다.

(4) 현장에서의 관찰: 특정 병동에서 갑작스러운 감염 사례 증가

- (의료관련감염) 사례: 특정 병동에서의 감염 사례 증가
- 상황: 의료진이 특정 병동에서 환자들 사이에서 감염 사례가 갑작스럽게 증가하는 양상을 관찰한다. 예를 들어, 동일 병동의 여러 환자에게서 비슷한 증상(발열, 상처 감염, 또는 호흡기 감염 등)이 단기간에 나타나는 경우, 이는 감염 확산의 초기 신호일 수 있다. 이러한 상황은 감염관리 담당자나 의료진이 현장에서 직접 확인하거나 일상적인 환자 관찰 중 발견될 수 있다.
- 보고 절차
  - 초기 확인: 감염 사례와 관련된 증상, 발생 시간, 환자 분포 등을 기록하여 감염 패턴을 확인한다.
  - 즉각 보고: 의료진은 관찰된 사례를 즉시 병원 감염관리실에 보고하여, 감염 가능성을 평가하도록 요청한다.
- 추가 조사 요청: 감염관리실은 보고된 사례에 대해 구체적인 데이터 분석과 사례정의(case definition)를 수행하여 감염 유행 여부를 확인한다.
- 역학조사 및 대응: 감염원으로 의심되는 병원 환경(병실 표면, 장비, 의료기기) 또는 특정 행위(수술, 치료 과정) 등을 조사한다. 필요시, 해당 병동의 환자와 의료진을 대상으로 선제적 감염검사를 시행하거나 격리 조치를 권고한다.

결과: 현장에서의 신속한 관찰과 보고는 감염 확산을 조기에 차단할 수 있는 핵심 단계로, 병동 내 환경 정비, 감염원 제거, 예방적 감염관리 강화 등을 통해 추가적인 피해를 줄일 수 있다.

3 역학조사의 단계와 방법

1. 사전 준비 및 초기 대응

의료관련감염 및 의료관련감염병의 유행 발생 시 신속한 대응과 체계적인 조사가 필요하다. 이를 위해 다음과 같은 단계가 수행된다.

1) 정보 수집

감염병의 초기 확산을 억제하고 적절한 대책을 마련하기 위해, 발생 상황과 관련된 기초 정보를 수집한다. 주요 정보에는 다음이 포함된다:

- 병원체 정보: 원인 병원균의 종류 및 특성.
- 감염 사례 수: 발생한 감염의 총 건수 및 환자 분포.
- 발생 위치: 감염이 집중적으로 발생한 병동 또는 구역.
- 환자 특성: 환자의 연령, 기저질환, 면역 상태 등.
- 진단 검사 정보: 실험실 검사 결과, 배양 검사 등.
- 초기 조치 여부: 이미 시행된 통제 조치 및 그 효과.

2) 대응 수준 결정

수집된 정보를 바탕으로 추가 조사가 필요한지, 혹은 병원 자체 조사로 충분한지를 판단한다. 대응 수준은 감염의 규모, 중증도, 그리고 감염병의 전파 가능성에 따라 결정된다. 필요 시 관할 보건소나 지자체에 협조를 요청할 수도 있다.

3) 초기 관리 조치

초기 통제 조치는 감염병의 확산을 신속히 억제하고, 추가 사례를 방지하며, 원인 규명을 위한 시간을 확보하는 데 중요한 역할을 한다. 아래는 실제 사례를 통한 초기 관리 조치의 예이다.

사례

(의료관련감염) 사례 1: 수술부위 감염(Surgical Site Infection, SSI)

- 상황: 특정 수술 후 다수의 환자에게서 감염 발생.
- 조치
  - 해당 수술 장비의 사용 중지 및 멸균 과정 점검.
  - 감염된 환자 치료 및 병원균 배양 검사 실시.
  - 수술 절차에 관여한 의료진과 환경에 대한 소독 강화.

(의료관련감염병) 사례 2: 중환자실에서 CRE 발생

- 상황: 중환자실에서 Carbapenem-Resistant *Enterobacterales* (CRE) 감염 사례 급증.
- 조치
  - 환자 코호트 격리 및 의료진 개인보호구(PPE) 착용 의무화.
  - 중환자실 내 의료기기 및 공용 물품의 철저한 소독.
  - 감염 환자의 이동 제한 및 병원 내 다른 병동으로의 전파 차단.

2. 역학조사 설계 및 데이터 수집

1) 진단의 정확성 확인

감염병 발생 초기 단계에서 유행(outbreak)이 실제인지 여부를 확인하여 불필요한 자원(인력과 시간등) 낭비를 줄이고 적절한 대응 조치를 마련한다.

먼저, 환자 기록과 병력을 검토하여 증상이 나타난 시간, 감염이 발생한 위치, 그리고 환자의 치료 내역을 분석한다. 이를 통해 감염이 전형적인 양상을 보이는지 평가하고 유행 가능성을 판단한다. 또한, 병리학적 소견과 임상 증상을 확인하여 질병의 전형적인 임상 양상이 나타나는지 분석한다. 이는 감염 사례가 동일한 병원체로 인한 것인지 여부를 평가하는 데 도움을 준다. 아울러, 의료진과의 인터뷰를 통해 진단 과정과 사용된 진단 방법의 적절성을 논의하여 진단의 신뢰성을 높인다.

다음으로, 실험실 오류 평가가 이루어진다. 사용된 검사 방법, 예를 들어 배양 검사나 PCR 검사가 적절하고 신뢰할 수 있는지 확인한다. 이 과정에서 샘플이 채취, 보관, 분석되는 동안 오염이 발생했을 가능성을 검토하며, 필요하다면 검사 결과를 반복 확인하여 동일한 샘플이나 새로운 샘플로 추가 검사를 실시한다.

감시 기준 변경과 사례정의 평가도 중요한 단계이다. 감시 기준이나 사례정의가 변경되었는지 확인하여, 새로운 기준으로 인해 사례 수가 증가했는지를 분석한다. 또한, 새로운 사례정의가 기존 데이터

와 얼마나 일치하는지 비교하여 감염 발생의 패턴을 재평가한다.

마지막으로, 임상 및 미생물학적 결과를 검토한다. 환자의 주요 증상과 진단된 질병이 일치하는지 분석하고, 병원체 배양 결과와 항생제 감수성 검사 결과를 통해 감염의 원인과 확산 가능성을 평가한다.

이와 같이, 진단을 정확히 확인하고, 실험실 오류 여부를 평가함으로써, 유행 발생을 명확히 판단할 수 있으며, 효율적인 역학조사를 시행할 수 있다.

의료관련감염 예시

- 특정 병동에서 동일한 병원체(예: *E. coli*)가 반복 검출되었으나 임상 증상이 일치하지 않는 경우.
- 조치: 실험실 오류 여부를 확인(검사 과정에서 오염 가능성)
  - 가유행(pseudo-outbreak) 여부 확인
  - 사례정의와 감시 기준을 재검토.
  - 추가 환자 데이터와 환경 샘플 채취를 통한 원인 파악

2) 대응팀 구성

의료관련감염병 또는 의료관련감염의 유행이 확인된 후, 효과적인 대응을 위해 적절한 이해관계자들과 협력 체계를 구축하는 것이 필수적이다. 이를 통해 감염 확산을 조기에 차단하고 문제를 해결할 수 있다.

먼저, 병원 내부에서는 다양한 관계자가 대응팀의 핵심 역할을 수행한다. 병원 책임자와 관련 의료진은 감염 상황의 전반적인 조정과 실행 계획의 승인을 담당하며, 감염관리실장은 발생 현황을 모니터링하고 필요한 조치를 주도적으로 이행한다. 감염관리간호사는 감염 사례를 기록하고 예방 조치를 실질적으로 수행하는 데 중요한 역할을 한다. 또한, 시설 환경 부서는 병원 내 위생을 유지하고 감염 발생 구역의 소독 및 정비를 담당한다. 홍보팀은 감염병 관련 정보를 의료기관 내부 및 외부에 명확히 전달하며, 실험실 담당자는 감염원 규명과 관련 데이터를 분석하여 과학적 기반을 제공한다.

이와 함께, 의료기관은 지자체 및 관련 부처와 긴밀히 협력해야 한다. 지자체는 역학조사를 주관하고 상황에 대한 지원과 조율을 담당하하고, 필요한 자원과 관리 계획 수립을 지원한다. 질병관리청은 국가 차원의 전문적 조언과 지침을 제공하며, 식품의약품안전처(FDA)는 의료기기 및 약물 관련 문제 발생 시 기술적 지원을 제공한다.

대응팀 구성 이후에는 철저한 계획과 현장 준비가 필요하다. 먼저, 누락된 정보를 확인하여 추가로 필요한 데이터를 식별하고, 이를 조사 계획에 포함한다. 현장 방문 계획을 수립하여 감염 발생 구역을 직접 평가하고 추가 정보를 수집한다. 이 과정에서 설문지, 기록용 장치, 분석 소프트웨어 등 데이터 수

집과 분석에 필요한 도구를 사전에 준비한다.

이와 같은 대응팀 구성 및 현장 준비는 감염병 유행에 신속하고 체계적으로 대응할 수 있는 기반을 제공한다. 이를 통해 감염 확산을 억제하고, 환자와 의료진의 감염으로부터의 안전을 보장하며, 병원의 정상적인 운영을 유지하는 데 기여할 수 있다.

3. 유행 조사(의료관련감염 및 의료관련감염병)

1) 유행 확인

유행 확인은 사례 수 증가, 진단 기준 변경, 그리고 진단 절차 오류 가능성을 검토하여 감염병이 실제로 확산되고 있는지 평가하는 단계이다. 먼저 감염 사례의 수가 과거 데이터와 비교하여 통상적인 발생률을 초과하는지 확인한다. 또한, 진단 기준의 변경으로 인해 사례 수가 증가한 것은 아닌지 평가하며, 진단 절차에서 오류가 발생했을 가능성도 검토한다. 이러한 검토 과정을 통해 유행 여부를 명확히 파악하고 불필요한 자원 낭비를 방지한다.

2) 사례정의 설정: 임상 및 실험실 기준, 위치, 시간 범위를 포함한 정의 개발

(1) 넓은 사례정의(Broad Case Definition)

초기 조사 단계에서는 가능한 모든 사례를 포함하도록 사례정의를 설계한다. 이를 통해 감염병 확산의 전체 범위를 파악하고, 추가 데이터를 수집하면서 점진적으로 구체화된 정의를 개발한다.

의료관련감염 사례정의의 작성 예시

- **Methicillin-resistant *Staphylococcus aureus* (MRSA) 사례정의 예시**

병원 A의 신생아 중환자실에 입원한 동안 MRSA 혈류감염이 발생한 사례(2024년 1월 1일~12월 31일).

- ***Burkholderia cepacia* 사례정의의 예시**

2024년 1월 1일부터 6월 30일 사이 약물 A를 받은 환자에서 유행 병원체와 일치하는 *B. cepacia* 배양 결과가 확인된 사례.

(2) 사례정의의 계층화

사례정의는 불확실성을 반영하여 나누어 볼 수 있다.

① 확진 사례(Confirmed): 실험실 검증이나 명확한 진단 결과로 확인된 사례를 포함한다. 예를 들어, MRSA 혈류감염이 실험실 검사로 확인된 경우이다.

② 추정 사례(Probable): 전형적인 임상 증상과 역학적 연관성이 있으나, 실험실 검증이 없는 사례를 의미한다. 예를 들어, 동일 병원체 감염 확진자와 접촉한 후 유사한 증상을 보이는 환자가 이에 해당한다.

③ 의심 사례(Possible): 전형적인 임상 증상이 부족하거나 역학적 연관성이 약한 경우로, 노출 이력은 있으나 증상이나 실험실 증거가 불충분한 사례가 포함된다.

3) 사례 식별 및 계산: 과거 및 현재 발생 사례를 체계적으로 확인

감염병 유행을 정확히 평가하기 위해 사례를 체계적으로 식별하고 분석하는 과정은 필수적이다. 이 단계에서는 과거와 현재 발생한 사례를 확인하고, 수집된 데이터를 기반으로 사례를 분류하여 추가 조사를 통해 확진 사례를 정리한다.

사례 식별은 다양한 자료원을 활용하여 이루어진다. 의료기기 사용 내역은 내시경, 카테터와 같은 특정 기기와 관련된 감염 사례를 확인하는 데 중요한 정보를 제공한다. 또한, 의료진의 감염 사례 및 노출 기록을 분석하여 의료진이 유행의 중심에 있는지 평가한다. 수술 기록과 병동 관리 조치 이력은 동일한 절차나 환경에서 발생한 사례를 확인하는 데 활용되며, 환자 이동 기록은 병원 내에서 감염이 어떻게 전파되었는지 파악하는 데 도움을 준다.

미생물학 보고서는 배양 검사 결과와 병원체 특성을 통해 감염의 원인을 구체적으로 밝히는 데 사용된다. 이와 함께 의무 기록, 의사 인터뷰, 영상 보고서, 그리고 항생제 처방 내역은 환자의 병력과 치료 과정을 종합적으로 검토하는 데 기여한다. 감시 기록과 특정 약품이나 의료 장비의 사용 내역은 감염 사례와 특정 행위 또는 약품 장비등과의 역학적 연관성을 평가하는 데 유용하다.

사례를 식별한 후에는 체계적인 분석 과정을 거쳐 데이터를 정리한다. 먼저 각 환자에게 고유 식별번호를 부여하고 증상, 검사 결과, 노출 위치(장소) 등 필수 정보를 포함한 라인 리스트를 작성한다. 이를 통해 수집된 데이터를 정리하며, 동일한 환자의 중복된 기록이나 잘못된 데이터는 제거한다. 누락된 데이터가 있을 경우 추가 인터뷰와 기록 검토를 통해 수정 및 보완한다.

발생률 계산은 사례 데이터의 시간적 변화를 분석하는 데 중요하다. 이를 일일 또는 주간 단위로 분석하여 사례 증가 추세를 파악하고, 전파 패턴을 이해하기 위해 유행곡선을 작성한다. 유행곡선은 유행의 시작, 정점, 그리고 감소 단계를 시각적으로 보여주어, 적절한 방역 조치를 결정하는 데 핵심적인 정보를 제공한다.

이와 같은 사례 식별 및 계산 과정은 유행의 정확한 범위를 파악하고, 감염 확산을 억제하기 위한 실

질적인 근거를 제공한다. 이는 감염병 유행의 기초 원인을 규명하고, 향후 예방 전략을 수립하는 데 필수적인 단계로 활용된다.

4) 데이터 수집 및 분석: 사례 정보 표 정리, 감염 분포를 설명하는 기술 역학 분석 수행

감염병 유행의 특성을 명확히 이해하고 효과적으로 대응하기 위해 데이터 수집과 분석 과정은 체계적으로 이루어져야 한다. 이 과정은 사례 데이터를 정리하고 분석하며, 감염병의 분포와 전파 패턴을 설명하는 데 초점이 맞춰진다.

우선, 사례 정보를 체계적으로 관리하기 위해 라인리스트를 작성한다. 이는 환자 정보를 표 형식으로 정리하여 데이터를 명확히 시각화하고 분석할 수 있도록 돕는 도구이다. 라인리스트에는 환자가 발생한 병원의 위치(병동, 병실 번호, 침대 번호 등)가 포함되며, 이를 통해 감염이 특정 구역에 집중되었는지 파악할 수 있다. 환자의 연령, 성별, 인종 등 인구학적 정보는 환자의 특성을 분석하는 데 활용되며, 주요 증상과 실험실 검사 결과와 같은 임상 정보는 감염병의 전형적인 양상을 평가하는 데 도움을 준다. 또한, 침습적 의료기기 사용이나 면역 저하 상태와 같은 위험 요인은 감염의 원인 및 전파 경로를 파악하는 데 중요한 역할을 한다.

데이터 수집과 정리는 표준화된 도구와 절차를 통해 이루어진다. 예를 들어, 표준화된 설문지를 활용하여 환자 인터뷰를 진행하거나 의료 기록을 검토하여 정보를 수집한다. 수집된 데이터는 스프레드시트와 같은 도구를 사용하여 정리하며, 열은 변수(예: 연령, 증상), 행은 각 사례를 나타내도록 구성한다. 이러한 방식은 데이터를 체계적으로 관리하고 분석의 효율성을 높이는 데 기여한다.

감염병 유행의 시간적 분포를 이해하기 위해 유행곡선을 작성한다. 유행곡선의 X축은 시간 단위를 나타내며, 일, 주, 또는 월 단위로 설정된다. Y축은 해당 시간 단위에 발생한 사례 수를 표시한다. 사례는 확진, 추정 등으로 분류되어 그래프에 나타나며, 이를 통해 각 유형의 사례가 어떻게 증가 또는 감소했는지 분석할 수 있다. 유행곡선은 유행의 시작, 정점, 그리고 감소 단계를 명확히 보여주며, 감염병 관리와 대응 전략 수립에 있어 핵심적인 도구로 활용된다.

이와 같은 데이터 수집 및 분석 과정은 감염병의 발생 규모와 분포를 명확히 파악하고, 감염 원인을 규명하며, 적절한 방역 및 예방 대책을 수립하는 데 중요한 기초를 제공한다. 또한, 체계적이고 시각화된 분석 결과는 의사결정자에게 명확한 정보를 제공하며, 신속하고 효과적인 감염병 대응을 가능하게 한다.

4. 관리 및 예방

1) 감염 관리 평가

감염 관리 평가 과정은 현장 방문을 통해 감염 관리 절차와 실제 관행을 분석하는 데 초점이 맞춰진다. 주요 활동으로는 약물 준비, 침습적 처치 관리, 손 위생, 격리 절차, 의료기기 소독 및 재처리, 환경 관리를 포함하여 의료진의 작업 방식을 관찰한다. 의료진 간 동일 절차를 수행하는 기술과 실행 방식의 차이를 확인하고, 지침 또는 매뉴얼 권고 사항과 실제 실행 간의 불일치 여부를 평가한다. 이러한 차이는 의료진의 교육 수준과 경험 차이에 의해 발생할 수 있다.

또한, 의료진과 논의를 통해 현재 감염관리 프로토콜의 최신화 여부 및 정확성을 평가하고, 프로토콜과 실행 간의 차이점을 문서화한다. 현장에서 프로토콜을 일관되게 적용하지 못하는 이유와 어려움도 분석하여 개선 방안을 모색한다. 이를 보완하기 위해 유사한 감염병 유행 사례와 최신 과학적 연구를 검토하고, 데이터를 체계적으로 정리하기 위한 추출 양식을 개발한다. 이를 통해 유행의 원인과 전파 경로를 구체적으로 이해하고 데이터를 표준화할 수 있다.

2) 환경 평가

환경 평가는 감염 경로와 관련된 환경적 요인을 분석하여 유행의 원인을 규명하는 과정이다. 환경 샘플링은 환자가 접촉한 표면, 의료기기, 물 공급, 공기 중 병원체를 대상으로 이루어지며, PCR과 같은 분자생물학적 검사 기법을 활용하여 환경 샘플과 임상 샘플 간의 역학적 연관성을 확인한다. 필요에 따라 공중보건 실험실의 도움을 받아 대량 샘플 분석을 진행한다(참고: 검체의 종류와 목적에 따라, 보건환경연구원, 식약처, 지방 식약청 등 검사 기관이 다름).

의료진 검사도 필요에 따라 수행된다. 특정 의료진과 환자 간의 감염 연관성이 의심될 경우, 환자의 병원체와 의료진 검사 결과를 비교하여 연관성을 평가한다. 이러한 검사 과정은 의료진에게 심리적 부담을 줄 수 있으므로 이를 사전에 고지하며, 검사 결과가 양성일 경우 감염원이 의료진인지 환자에게서 전파된 것인지 구체적으로 분석한다.

3) 통제 조치 권고

감염병 통제를 위해 병원체, 전파 경로, 환자 상태를 고려한 즉각적인 조치와 예방적 감염 관리 방안을 제안한다. 손 위생 강화와 의료진 대상 재교육, 격리 및 코호트 병동 운영, 병동 이동 제한과 출입 통제 등은 즉각적인 조치로 시행된다. 소독 및 멸균이 미흡한 의료기기의 사용을 중단하고, 장비 재처리 절차를 점검 및 개선하며, 고위험 구역의 소독 빈도를 증가시킨다. 예방적 조치로는 최신 감염관리 지

침에 기반한 교육과 병원의 프로토콜 강화, 유행 장소의 감염률 지속 모니터링, 신규 사례에 대한 신속 대응 체계를 구축한다.

4) 결과 해석 및 유행 모니터링

조사 결과를 종합하여 감염 통제 조치의 효과를 평가하고, 추가 사례 발생을 감시하는 체계를 유지한다. 분석 연구는 사례-대조군연구(Case-Control Study)와 코호트연구(Cohort Study)를 활용한다. 사례-대조군연구는 감염된 사례군과 감염되지 않은 대조군 간의 위험 요인 노출 빈도를 비교하며, 대조군은 사례군과 주요 특성이 유사해야 한다. 코호트연구는 특정 위험 요인에 노출된 그룹과 그렇지 않은 그룹 간의 감염 발생률을 비교하며, 유행이 특정 병동이나 인구에 국한된 경우 적합하다.

5. 사후 활동

감염병 유행이 통제된 이후에는 조사 결과를 정리하고, 조치의 효과를 평가하며, 유사한 상황이 재발하지 않도록 예방 대책을 마련하는 사후 활동이 이루어진다. 이 과정은 보고서 작성, 사후 평가, 그리고 정책 제안의 세 단계로 구성된다.

1) 보고서 작성 및 공유

유행 조사가 완료되면 최종 보고서를 작성하여 주요 이해관계자들과 공유한다. 이 보고서는 유행의 배경, 조사 방법, 결과, 권고사항 등을 포함하며, 유행의 전반적인 내용을 체계적으로 기록한다.

우선, 초기 데이터를 바탕으로 사례정의를 정제한다. 이를 통해 감염 사례를 보다 구체적으로 분류하고 분석력을 높이며, 필요한 경우 추가 데이터를 수집하여 사례정의의 민감도와 특이도를 최적화한다. 이후 정제된 사례정의를 활용하여 새로운 사례를 탐색하고, 기존 감염 발생이 종료되었는지를 확인한다. 유행 종료 여부를 평가하기 위해 최소 1개월에서 1년까지 감시 기간을 설정하여 지속적인 모니터링을 실시한다.

조사 결과는 구두 브리핑 형태로 지자체, 병원 관리자, 감염관리팀과 공유된다. 이 과정에서 유행의 주요 원인과 특성, 감염 관리 문제점, 그리고 통제 조치의 효과 및 향후 계획이 논의된다. 최종 보고서는 질병관리청 및 지자체에 제출되며, 유행 기록이 보존되고 후속 조치 계획이 수립된다. 보고서에는 유행의 특성, 감염관리 문제점, 통제 조치와 그 결과, 권고사항, 그리고 선택적으로 실험실 확인 결과가 포함된다.

2) 사후 평가

조사와 통제 조치가 성공적으로 이루어졌는지 평가하고, 감염 관리 체계를 검토하고 개선점을 도출하는 단계이다.

사후 평가의 주요 목적은 유행 통제의 효과를 확인하고, 감염 관리 체계의 취약점을 파악하며, 미래 대응 전략 설계를 위한 데이터를 도출하는 것이다. 평가 과정에서는 감염 발생률 감소 여부를 확인하고, 사례정의와 감시 체계의 효과를 분석한다. 또한, 의료진과 직원들이 감염관리 지침을 얼마나 준수했는지를 평가하며, 조사와 통제 과정에서 사용된 자원(인력, 장비, 시간)이 적절하게 활용되었는지도 검토한다. 다학제적 팀 간 협력과 의사소통의 효과를 분석하여 조직적 대응 능력을 평가한다.

조사 과정에서 발견된 문제점과 개선 사항은 의료진과 주요 이해관계자들에게 환류(Feedback)된다. 이 과정을 통해 향후 감염병 유행에 대한 대응 능력을 강화하고, 유사한 문제가 재발하지 않도록 예방 조치를 설계한다.

3) 정책 제안

사후 활동의 마지막 단계는 감염 관리 정책을 강화하고, 유사한 유행을 방지하기 위한 표준화된 지침을 마련하는 것이다.

정책 제안의 목표는 감염병 발생 예방 및 관리 능력을 강화하고, 병원과 공중보건 시스템 간 협력을 통해 지속 가능한 감염 관리 체계를 구축하는 데 있다. 이를 위해 조사를 통해 수집된 데이터를 기반으로 주요 원인과 문제점을 분석하고, 구체적인 권고사항을 도출한다. 이 권고사항은 새로운 감염관리 절차와 교육 프로그램 개발, 그리고 실행 가능한 정책으로 구체화된다. 또한, 정책의 비용과 효과를 분석하여 실현 가능성을 평가하고 우선순위를 설정한다.

참고 문헌

1. Christensen BE, Fagan RP. Chapter 18, Healthcare Settings. In: Rasmussen SA, Goodman RA, editors. The CDC Field Epidemiology Manual. New York: Oxford University Press; 2019.
2. Dixon RE. Control of Health-Care-Associated Infections, 1961-2011. MMWR Suppl. 2011;60(4):58-63.
3. Florida Department of Health. Conducting Health Care-Associated Infection Outbreak Investigations [Internet]. Tallahassee: Florida Department of Health; 2021 [Cited 2025 Jan 17]. Available from: https://www.floridahealth.gov/diseases-and-conditions/health-care-associated-infections/_documents/hai-guide-investigations.pdf.
4. Haque M, Sartelli M, McKimm J, et al. Health care-associated infections – an overview. Infect Drug Resist. 2018;11:2321-33.
5. Lumley SF, Constantinides B, Sanderson N, et al. Epidemiological data and genome sequencing reveals that nos-

ocomial transmission of SARS-CoV-2 is underestimated and mostly mediated by a small number of highly infectious individuals. J Infect. 2021;83(4):473-82.

6. OYong K, Coelho L, Bancroft E, et al. Health Care-Associated Infection Outbreak Investigations in Outpatient Settings, Los Angeles County, California, USA, 2000-2012. Emerg Infect Dis. 2015;21(8):1317-21.
7. World Health Organization. Principles and steps of an outbreak investigation [Internet]. Geneva: WHO; 2022 [cited 2025 Jan 17]. Available from: https://www.who.int/docs/librariesprovider2/default-document-library/module_2.1_outbreak-investigation_ds_220830.pdf.
8. World Health Organization. Surveillance of health care-associated infections at national and facility levels. 2024.

역학조사관 교육
표준교재

Textbook For
Korea Field Epidemiology
Training Program





7장

검역 및 해외유입 감염병의 관리







제1절 검역의 역할과 방법

제2절 해외유입 감염병의 예방과 관리

집필자 이윤현, 천병철
검토자 김승혁, 김옥수, 박진욱, 유효순, 정연화

제1절 검역의 역할과 방법

1 검역의 역할

1. 개요

1) 정의

검역은 해외에서 발생한 감염병이 국내로 유입되는 것을 방지하기 위해 공항, 항구 등에서 사람, 운송수단, 화물 등을 검사하고 필요한 경우 격리하거나 소독하는 공중보건 활동이다. 우리나라 검역법에서는 검역에 대해 우리나라로 들어오거나 외국으로 나가는 사람, 운송수단 및 화물을 검역(檢疫)하는 절차와 감염병을 예방하기 위한 조치에 관한 사항을 규정하여 국내외로 감염병이 확산하는 것을 방지함으로써 국민의 건강을 유지·보호하는 것을 목적으로 한다고 규정한다. WHO의 국제보건규칙(IHR, 2005)에서 검역(quarantine)은 감염이나 오염의 가능성이 있을 때 확산을 막기 위해 아프지는 않지만 감염이 의심되는 사람의 활동을 제안하거나 의심되는 수하물, 컨테이너, 운송수단 또는 물품을 다른 것과 분리하는 것을 말한다("quarantine" means the restriction of activities and/or separation from others of suspect persons who are not ill or of suspect baggage, containers, conveyances or goods in such a manner as to prevent the possible spread of infection or contamination). 미국 CDC에서 격리와 검역은 감염성 질환에 걸렸거나 감염 가능성이 있는 사람들과의 접촉을 차단하여 공중 보건을 보호하는 역할이라고 정의한다. 구체적으로 **격리**(isolation)는 감염성 질환이 확인된 환자를 건강한 사람들과 분리하여 질병 확산을 방지하는 조치(Isolation separates sick people with a quarantinable communicable disease from people who are not sick)이고, **검역**(quarantine)은 감염병에 노출된 사람들을 일정 기간 동안 분리하고 이동을 제한하여 감염 여부를 확인하는 조치(Quarantine separates and restricts the movement of people who were exposed to a contagious disease to see if they become sick)라고 정의한다. 격리와 검역은 단순한 보건 조치뿐만 아니라, 사회 전체의 이익을 위해 개인에게 자유를 제한할 수 있는 국가의 권한에서 비롯된 경

찰권(police power)의 기능도 수행한다.

2) 역사적 의미

'검역(quarantine)'이라는 용어는 라틴어로 40일을 의미하는 'Quaranta'에서 유래한다. 검역의 개념은 14세기 유럽에서 발생한 흑사병(페스트)과 밀접한 관련이 있다. 흑사병은 14세기 이탈리아에서 1347년부터 1351년 사이에 유럽 인구의 약 30~60%를 감소시킬 정도로 치명적인 감염병으로, 병원체인 예르시니아 페스티스(*Yersinia pestis*) 박테리아에 의해 발생했다. 이 병은 쥐 벼룩을 매개체로 하여 급속도로 확산되었고, 그 결과로 유럽 전역에 걸쳐 대규모 사망자가 발생했다. 검역은 이 흑사병의 확산을 막기 위해 도입되었고 당시 베니스로 입항하는 선박은 항구에 접안하기 전에 40일 동안 억류되었으며, 이 기간 동안 감염병에 감염되지 않았음이 확인된 후에야 항구로 들어올 수 있었다. 당시 사람들은 감염병의 원인을 명확히 이해하지 못했지만, 일정 기간 동안의 격리가 병의 확산을 막는 데 효과적이라는 것을 경험적으로 알게 되었다. 선원과 승객은 선박에서 내리지 못하게 했고, 이 기간 동안 감염병 증상이 나타나지 않으면 비로소 항구로 입항할 수 있도록 했다. 종교적인 관점에서 중세 유럽에서는 감염병을 신의 벌이나 죄의 결과로 여기는 경향이 있었으나 흑사병의 반복적인 대유행은 단순히 신앙적 접근으로 해결되지 않았고, 위생과 격리의 필요성이 점차적으로 받아들여졌고 결과적으로 흑사병이 공중보건의 개념을 발전시키는 계기가 되었다. 도시국가의 역할 베니스뿐만 아니라 제노바와 마르세유 같은 도시국가들도 검역소(lazaretto)를 설립하여 입항 선박, 화물, 승객을 격리했고 이러한 검역소는 감염병 확산을 줄이는 데 효과적이었다.

2. 검역체계의 표준화

흑사병 이후 공중보건이 강조되면서 도시들은 위생 상태를 개선하기 위해 노력했으며, 검역소와 같은 보건 인프라가 점차적으로 확대되었다. 현대에 이르러 국제적으로도 검역 체계를 표준화하려는 시도가 이루어졌고, 이는 WHO의 국제보건규칙(2005)으로 이어졌다. 이처럼 흑사병은 검역의 기원으로서 중요한 역사적 전환점을 제공하였고 단순한 격리를 넘어 현대 공중보건의 핵심 원칙인 질병 예방과 통제로 발전하여 오늘날 검역은 코로나19와 같은 신종 감염병 확산을 막는 중요한 도구로 여전히 사용되고 있다.

3. 패러다임의 전환

검역의 패러다임은 전통적인 "격리와 감시" 중심의 소극적 대응에서 보다 능동적이고 통합적인 공중보건 개입으로 변화하고 있고, 이는 감염병 예방 및 통제의 효율성을 높이고, 국제적 협력을 강화하려는 방향으로 발전하고 있다.

표 7-1-1. 검역의 패러다임 변화

구분	현재(As is)		개선(To be)
검역개념	• Regulation and isolation quarantine • Passive forms of surveillance • Supplier-centric quarantine • Single-layer quarantine • CIQ independent operation	➡	• Intelligence analysis quarantine • Active forms of surveillance • Demand-driven quarantine • Multi-layer quarantine • CIQ integrated operations
법 제도	• 경직된 검역체계 • 입국자 전수 검역체계	➡	• IHR의 PoE역량 고도화 위한 검역체계 • 유연한 검역체계 • 위험분석에 따른 핀세 및 타겟 검역체계
검역영역	• 유증상자 여행객 중심 검역 • 관능검사 중심 검역 • 공·항만 중심 검역	➡	• TGS 등 자발적 참여형 검역 • ICT 활용한 스마트 검역 • 지역 및 국제사회와 네트워크 검역

1) 현재(As-Is)

검역 개념에서 검역은 감염병 유입을 막기 위해 격리와 감시를 중심으로 소극적이고 제한적인 방식으로 이루어지고 있고 주로 승객을 대상으로 응급 대응에 초점을 맞추고 있다.

법적·제도적 체계에서 검역은 검역법에 근거하나 입국자 중심의 검역방식으로 운영되고 있다.

검역 영역에서 검역은 공항만과 승객 중심의 검역이 이루어지며, 지엽적 화물 중심의 관리로 인해 일부 검역 사각지대가 존재한다.

2) 개선(To-Be)

검역 개념에서는 감염병 예방을 위한 공중보건 개입으로 확대되며, 위험 평가를 기반으로 선별적 검사를 수행하고, 승객뿐만 아니라 화물과 운송수단에 대한 포괄적인 관리로 변화한다.

법적·제도적 체계에서 검역은 국제보건규칙을 기반으로 검역 관리의 핵심 역량이 강화되며, 여행자 건강 정보 제공 등 유연한 검역 체계로 전환되고 또한, 위험 분석에 따라 보다 전략적으로 선별적인 검

역이 이루어질 것이다.

검역 영역에서는 기존의 승객 중심 검역에서 벗어나 ICT를 활용한 국제물류, 국제 우편, 수입 농수산물 등 다양한 분야로 확대하여 검역 사각지대가 해소되며, 여행자에게 편익을 도모하는 자발적인 검역체계인 여행자 게놈 감시 프로그램(TGS)을 도입하여 검역의 효율성과 정확성이 증대될 것이다.

4. 검역의 미래

1) 기술 발전과 디지털 전환

검역 시스템은 디지털 기술과 빅데이터를 활용하여 더욱 정교화될 전망이다. 특히, 다음과 같은 기술이 검역에 도입될 것으로 예상된다. AI 기반 위험 평가가 가능해 입국자의 건강 정보를 사전에 분석하여 감염 가능성을 예측할 수 있을 것이고, 스마트 검역을 통한 열화상 카메라, 자동 소독 시스템 등 비접촉 방식의 검역 기술 확대될 가능성이 높다. 특히 빅데이터 통합 플랫폼 구축으로 글로벌 감염병 데이터 공유와 분석을 통해 실시간 대응 체계가 구축되어 검역의 정보화가 실현될 것으로 보인다. WHO는 감염병 유행 신속탐지, 실시간 경보제공, 공중보건 위기 신속 대응을 위해 관련 기관 간의 정보공유 체계인 글로벌 유행경보 및 대응네트워크(Global Outbreak Alert & Response Network, GOARN) 구축·운영 중인데 정보통신 및 인공지능 기술을 활용되고 있다.

2) 국제협력과 검역

글로벌화와 함께 감염병의 국경 간 확산 가능성이 커지면서 국제적인 협력이 필수적이다. IHR에서 국가 간 검역 협력의 중요성이 강조되고 있고 세계보건기구(WHO)와의 협력을 통한 국제 표준 준수사항도 마련되고 있어 향후 양자 간 혹은 다국 간 검역 훈련으로 감염병 대응 역량 강화를 위한 공동 훈련과 시뮬레이션이 이루어질 것이다. 이같이 글로벌화로 인해 국가 간 이동이 활발해지면서 감염병의 국제적 확산 위험이 증가하고 있어 검역은 이러한 위험을 최소화하고 국민의 건강을 보호하는 데 필수적인 역할을 한다. 특히 코로나19 팬데믹과 같은 상황에서 국제협력을 기반으로 하는 검역의 중요성은 더욱 부각되고 있다. 세계보건기구(WHO) 감염병 글로벌 위기 대응을 위해 실시간 질병감시와 정보 수집, 보고 및 대응 전 과정에서 글로벌 헬스 거버넌스를 재조직했다. WHO의 National Action Plan for Health Security (NAPHS)는 IHR 핵심역량의 구현을 가속화하는 국가 다년 기본계획에 검역을 적용하고 있다. 특히 글로벌보건안보구상(GHSA)은 개발도상국의 검사·감시·검역 능력 향상 지원, FAO(유엔식량농업기구) 및 OIE(세계동물보건기구) 등과 대응연계하고 있다. 검역은 감염병의 국내 유입을 방지하고 국민의 건강을 지키기 위한 중요한 공중보건 활동으로써 국제적인 협력과 철저한 검역

절차를 통해 감염병의 확산을 효과적으로 막을 수 있다.

3) 지역사회 연계형 검역

향후 검역은 지역사회 연계라는 측면에서 지역 보건소와 병원 등과의 협력을 통해 모니터링 체계를 구축하고 통합형 공중보건 체계 속에서 독립적인 활동이 아닌 공중보건 체계의 한 요소로서 작동할 것이며 이는 예방 중심의 접근으로 감염병이 발생하기 전에 위험 요인을 차단하는 예방적 활동이 강화될 것으로 보인다. 미래에는 검역이 공항만에 출입국자 감염병 감시를 주된 임무로 한정하지 않고 지역사회 방역과 연계를 통한 다분야 협력적 검역시스템을 구축할 것이다. 우리나라는 질병관리청 산하 5개 권역질병대응센터(수도권, 충남권, 호남권, 경북권, 경남권) 및 제주출장소를 설치하여 지역방역을 지원하고 있고 지역사회 자체대응역량을 강화하기 위해서 초동대응 및 위기 시 신속한 대응을 위해 시 도 감염병관리지원단 설치했다.

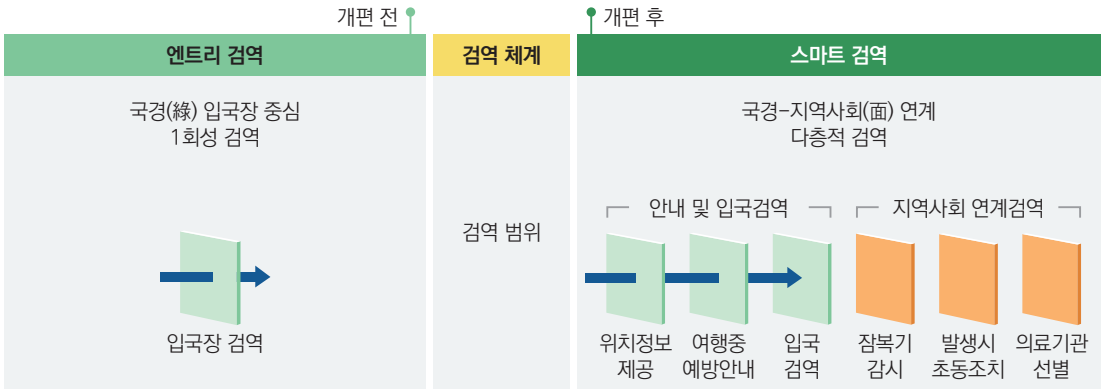


그림 7-1-1. 검역의 패러다임 변화

4) 신종 감염병의 대응

검역에서 가장 중요한 것은 새로운 감염병에 대한 대응이다. 코로나19와 같은 신종 감염병의 등장으로, 보다 유연하고 신속한 검역 체계가 요구되며 변종 바이러스 대응 전략을 수립하여 백신, 치료제와 함께 검역 단계에서의 초기 차단을 강화하는 방안을 마련하여야 할 것이다. 또한 국내외 유입 감염병 통합 관리라는 측면에서 국내외 발생 감염병 정보를 통합 관리하고 이에 맞는 검역 지침이 개발되어야 할 것이다. 질병관리청은 다음 팬데믹을 대비하여 국외에서 발생하거나 유행할 가능성이 있는 국외 감염병 정보를 수집·분석·관리하기 위한 전문인력을 확보하고 감염병 데이터의 질적인 제고와 신뢰성을 높은 자료를 생산하여 이를 연계한 해외 유입 감염병 대응전략을 수립해야 한다. 이를 위해 병원체

분리 및 유전체 진단능력의 혁신적 발달이 필요하며 분자탐지기술(실시간 PCR 검사), 휴대용진단기기, 실험실 검사키트 등 신속히 미생물을 분리하거나 염색체를 검사하는 기술로 발전하고 있다.

5) 사회경제적 파급 최소화

코로나19는 최악의 경기침체를 가져와 경제 사회 전반에 미치는 파급효과의 규모와 깊이를 알 수 없다. 그림 7-1-2에서와 같이 2020년 1~2분기 사이 큰 폭의 GDP 하락을 보였고, 이후 계속되는 추경으로 가까스로 GDP가 회복된 것처럼 감염병의 확산이 사회경제에 미치는 영향이 매우 크다. 지난 20년 동안 우리나라에 영향을 크게 준 Epidemic급 이상의 감염병은 2002년 SARS, 2015년과 2018년 메르스, 2020년 COVID-19, 4회 정도였으며, 선제적으로 검역단계에서 완전 차단에 성공하였다면 2018년 메르스 기준 약 10조 8천억 원 절감이 가능할 수 있었고, 이 중 50%만 검역으로 차단하게 되었다면, 약 5조 원 수준의 경제 절감 효과가 있었을 것이다. 결과적으로 검역은 해외로부터 유입되는 감염병을 효과적으로 차단할 수 있도록 단순히 감염자를 격리하는 소극적 방식에서 벗어나, 공중보건 전반을 아우르는 통합적이고 능동적인 방향으로 발전해야 하고, 이는 기술적, 법적, 제도적 개선과 국제 협력 강화로 국가 경제적 파급효과를 최소화하는 데 기여할 것이다.

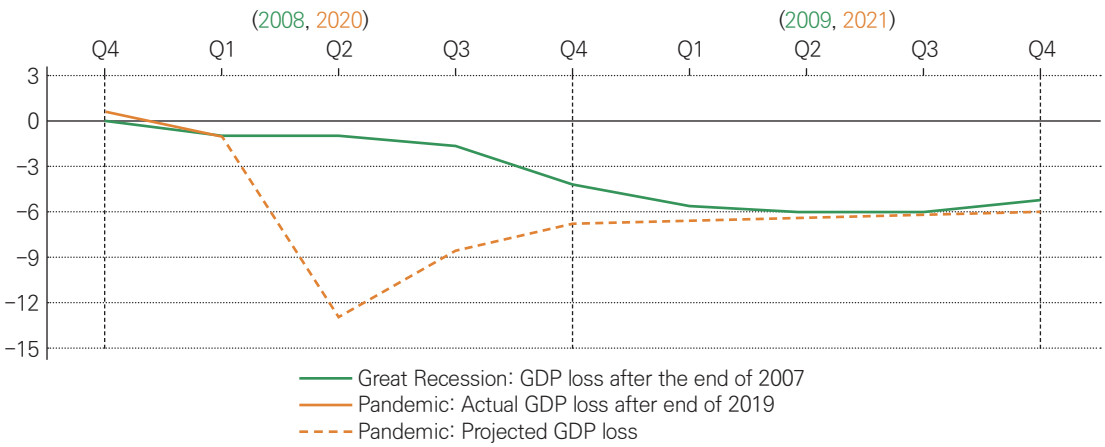


그림 7-1-2. COVID-19로 인한 GDP 감소; GDP % 차이

2 검역의 기능

1. 입국자 및 운송수단 검사

1) 입국자의 검사

WHO는 "입국지점(point of entry)"에서 검역은 여행객, 수하물, 화물, 컨테이너, 운송수단, 상품 및 우편 소포의 국제 입국 또는 출국 통로와 입국 또는 출국 시 이들에게 서비스를 제공하는 기관 및 지역에서의 활동을 강조하고 있다. 우리나라의 경우 코로나19 팬데믹 동안 국내 입국지점(point of entry) 검역 과정에서 해외 유입환자의 43%를 screening 하였고 유증상자 대상으로 COVID-19 진단검사를 공항검역소에서 약 2만 건(2020년 기준)을 수행하여 해외 감염병의 지역사회 유입을 최소화함으로써 검역의 중요성이 크게 주목받았다.

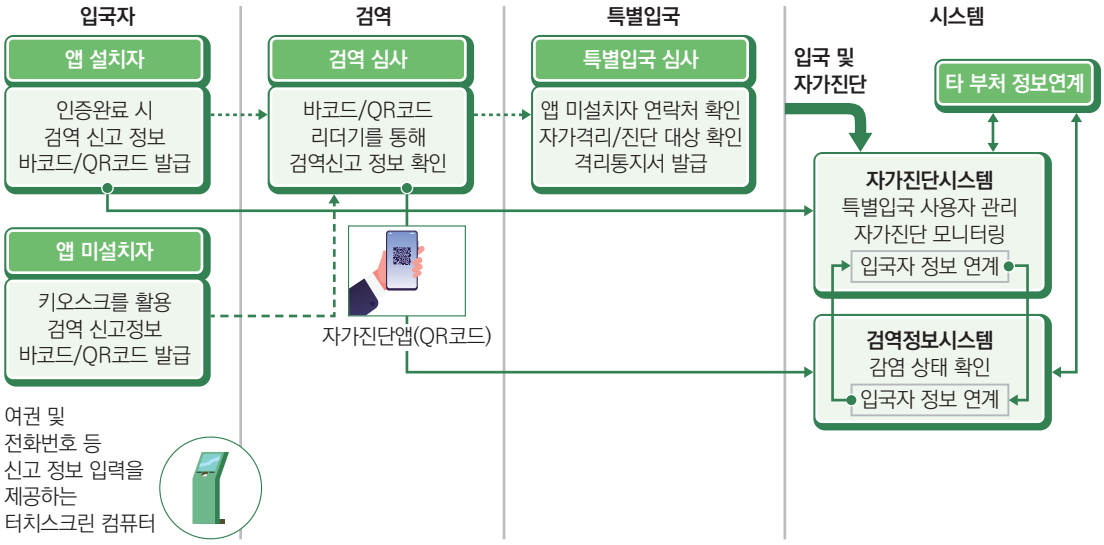


그림 7-1-3. 코로나19 모바일앱을 활용한 스마트 검역시스템

특히 COVID-19 이후 검역환경이 급격히 변화됨에 따라 특별입국 절차, 특정 지역 입국자에 대한 전수진단검사실시, 능동모니터링, 전 세계 입국자에 대한 격리와 감시 기간 내 진단검사, 자가격리, 임시생활시설, 공무 등 격리면제직종, 검역정보시스템, 감염병 재난위기관리 표준 매뉴얼(보건복지부) 및 실무 매뉴얼(각 부처) 운영 등 검역기법이 다양하게 진화되었다. WHO는 각종 감염병과 재난적 상황에

대한 신고, 보고를 대폭 강화하여 위기평가 알고리즘에 따라 24시간 안에 보고하도록 하고, 직접 현장 확인을 하거나 국가 이외의 조직이나 단체를 접촉하여 자료를 요구하는 권한을 가지고 물자와 여행 금지 등의 강력한 조처를 하였고, 각국에 자원 보강과 제도개선을 요구하고 있다. 그러나 이를 정비한 나라는 194개 회원국 중 20% 내외에 그치는 것으로 확인되었고, 문제해결의 일차적인 책임이 있는 감염병 발생 당사국은 오히려 의료시설의 확충, 개인보호장비와 필요 약품을 조달받아야 하는 상황이며 의료진 동원을 위한 예산 부족과 조직의 한계로 인하여 WHO 중심의 전 세계 국가들이 굳건한 공조체계 보완이 절실한 상황이다.

2) 운송수단의 검사

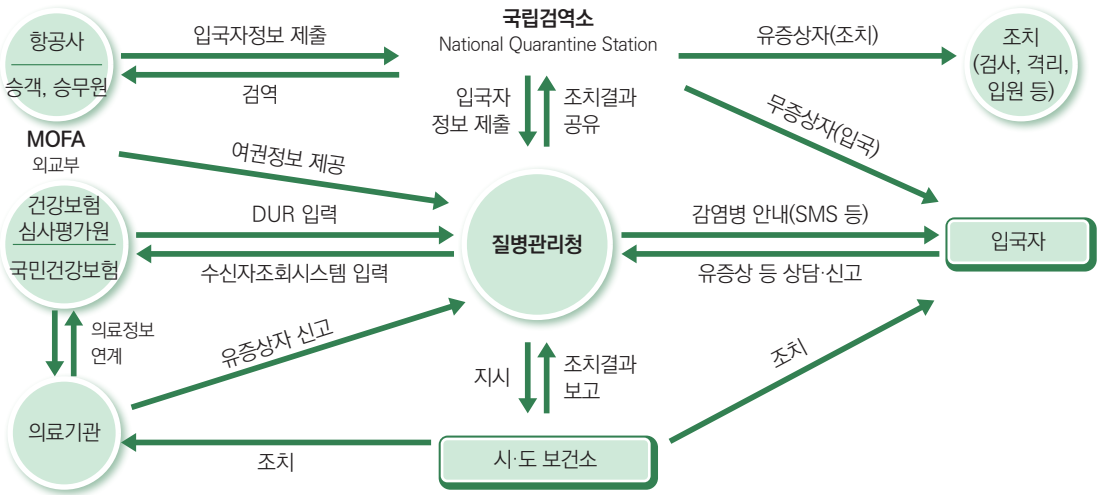


그림 7-1-4. 우리나라의 검역 정보화 체계

우리나라 검역법에 운송수단은 선박, 항공기, 열차 또는 자동차를 말한다고 정의한다. 우리나라로 들어오거나 외국으로 나가는 운송수단에 대한 검역은 운송수단 및 화물의 보건·위생 상태에 대한 경과(經過)와 현황, 출입국자의 검역감염병 감염·위험요인 여부 및 예방관리에 관한 사항, 운송수단의 식품 보관 상태, 감염병 매개체의 서식 유무와 번식 상태에 대해 조사하도록 하고 있다. 검역조사 결과, 감염병 매개체가 서식하거나 서식하는 것으로 의심되는 운송수단과 화물을 소독하고 감염병 매개체를 없애도록 운송수단의 장이나 화물의 소유자 또는 관리자에게 명하도록 하고 있다. 또한 검역소장은 선장 또는 선박의 소유자가 선박위생 증명서 발급을 신청하면 그 선박에 대하여 검역감염병 병원체의 오염 여부와 감염병 매개체 유무 등에 관한 조사를 하고, 그 결과 해당 선박에 검역감염병 병원체의 오염 의심이

없고 감염병 매개체가 서식하지 아니한 경우에는 6개월간 유효한 선박위생관리 면제증명서를 내주는 업무를 한다.

2. 검역감염병의 역학조사

1) 검역감염병

우리나라의 검역감염병은 '중동호흡기증후군(MERS), 중증급성호흡기증후군(SARS), 동물인플루엔자 인체감염증, 신종인플루엔자, 에볼라바이러스병, 콜레라, 페스트, 황열과 감염병으로서 외국에서 발생하여 국내로 유입될 우려가 있거나 국내에서 발생하여 국외로 전파될 우려가 있어 질병관리청장이 긴급 검역조치가 필요하다고 인정하여 고시하는 감염병'으로 정의하고 있다. 일본은 에볼라바이러스병, 크리미아콩고출혈열, 두창, 페스트, 마버그열, 남아메리카출혈열, 라싸열 등 15개 감염병을 검역감염병으로 지정하고 있다. 미국은 COVID-19, 에볼라바이러스병, 인플루엔자 등 호흡기 관련 질환을 중심으로 검역 중점 질병을 관리 운영하고 있다. 세계보건기구는 국제보건규칙을 통해 국제적으로 우려되는 공중보건 비상사태를 선언할 수 있으며, 이에 따라 특정 감염병에 대한 검역 조치를 권고한다. 이러한 감염병에는 에볼라바이러스병, 중증급성호흡기증후군(SARS), 중동호흡기증후군(MERS), 신종인플루엔자 등이 포함된다. 이러한 감염병들은 국제 여행과 무역을 통해 전파될 위험이 높아, 각국의 검역 과정에서 특별한 주의와 관리가 필요하다.

2) 역학조사

검역에서 역학조사는 검역구역 내에서 감염병에 관한 역학조사를 수행하여 감염병의 유입 및 확산을 방지한다. 우리나라는 공·항만에서 감염병 확진자가 발생할 경우 검역관이 역학조사관, 방역담당자 등과 협력하여 역학조사 및 검역조치를 신속하게 수행할 수 있는 체계를 갖추고 있다. 미국은 CDC 항만 검역소(Port Health Station)에 공중보건 요원, 검역 의료담당자 및 기타 인력을 배치하여 감염병의 유입 위험을 탐지, 대응 및 완화하기 위해 노력하고 있다. 검역소의 공중보건 요원은 여행자, 화물 및 운송수단(선박, 비행기)을 검사하여 질병의 징후를 확인하고 감염병의 위험요인에 노출되었을 가능성이 있는 여행자를 평가하는데, 여기에는 신체검사, 증상 확인, 추가 진단검사 및 역학조사를 수행한다. 그 결과, 감염병이 의심되는 여행자에 대해 CDC는 해당 개인을 격리하거나 역학조사를 수행할 수 있는 권한으로, 공중보건에 위협이 되지 않을 때까지 통제된 환경에서 그들을 구금할 수도 있다. 미국에 입국하는 여행자는 공중보건 모니터링의 일환으로 연락처 정보를 제공하라는 요청을 받는 경우가 많다. 특히 팬데믹과 같이 감염병 확산이 심한 경우 이러한 정보가 접촉자 추적이나 역학조사시 사용될

수 있으며, 여행자가 감염병에 걸린 것으로 확인되면 이를 통해 접촉자를 추적할 수 있다.

3. 여행건강정보 제공 및 자발적 참여

1) 여행자 건강정보

우리나라는 해외감염병 NOW를 통해서 여행자에게 방문하는 국가에 대한 감염병 및 안전정보를 제공하고 있고 향후 여행건강알림e 플랫폼으로 개선하여 정보 활용도를 높이기 위해 미국의 Travelers' Health Yellow Book, 영국의 Travel Health Pro 등의 주요 국가 여행건강정보와 연계하고, 특히, 감염병 발생으로 여행 중 감염될 위험이 높은 검역관리지역을 여행하는 여행자에게 감염병 정보를 알림톡, SMS로 제공하는 방안을 검토 중에 있다. 캐나다는 2018년 1월 세계경제포럼(WEF)에서 발표된 'Known Traveller Digital Identity(안전한 여행자 디지털 확인)' 프로그램을 네덜란드와 공동으로 시범 운영을 결정하였고 이 앱(App)을 통해 여행자 정보를 입력 및 출입국심사 당국의 정보공유 동의 절차를 사전에 거치면 입국심사 시간을 단축할 수 있게 되며, 해당 여행자의 여행기록이 축적될수록 'Known Traveller(안전한 여행자)'로 분류되어 향후 특별심사라인을 통해 더욱 빠르고 편리한 출입국심사를 받게 되는 구조이다. 일본 검역소는 해외 감염증의 유행 상황이 매일 변화하는 상황을 신속히 수집하여 감염증의 역학, 진단, 치료 등에 관한 새로운 정보를 가능한 한 알기 쉽게 제공함으로써 여행객이 안전한 여행을 즐길 수 있도록 한다.

2) 자발적 검역 참여

(1) 미국

자발적인 검역 참여 수단으로 여행자 기반 유전자 감시 시스템(TGS)을 도입하여 정보검역의 서막을 열었다. 유전자 감시 시스템은 질병통제예방센터(CDC)가 관리하는 더 넓은 생물보안 이니셔티브의 일환으로 미국 주요 공항에 도착하는 국제 여행자로부터 유전자 데이터를 수집하여 새로운 병원체, 특히 SARS-CoV-2의 새로운 변이와 인플루엔자 및 RSV 같은 호흡기 바이러스를 감시하고 탐지한다. 이 시스템은 2021년 몇몇 공항에서 파일럿 프로그램으로 시작되었으며, 이후 로스앤젤레스, 샌프란시스코, 뉴욕과 같은 주요 국제 허브를 포함한 최대 20개의 검역소로 확장되었다. 미국의 TGS 시스템은 잠재적인 발병에 대한 조기 경고를 제공하여 공중보건 관계자들이 더욱 효과적으로 감시하고 대응할 수 있도록 하고 특히, 이 프로그램의 새로운 바이러스 변종 및 박테리아 탐지 능력은 미국의 지속적인 팬데믹 대비에 필수적이다.

여행자 기반 유전자 감시체계(TBGS)

8 key points summarizing the Traveler-Based Genomic Surveillance (TGS) program:

1. Traveler Participation: Over 360,000 travelers from 15 countries have participated.
2. Airports: The program expanded from 3 to 9 airports, including JFK, LAX, and Miami.
   - Initial Airports: Launched at JFK, Newark, and SFO.
3. John F. Kennedy International Airport (JFK) 2. Newark Liberty International Airport (EWR) 3. San Francisco International Airport (SFO) 4. Hartsfield-Jackson Atlanta International Airport (ATL) 5. Los Angeles International Airport (LAX) 6. Miami International Airport (MIA) 7. Chicago O'Hare International Airport (ORD) 8. Washington Dulles International Airport (IAD) 9. Seattle-Tacoma International Airport (SEA)
4. Early Detection: TBGS detected the BA.2 variant 7 days early and BA.3 variant 43 days early
5. Funding: TBGS is part of the $1.75 billion allocated under the American Rescue Plan Nasal Swabs: Samples collected via voluntary nasal swabs
6. Monitoring Pathogens: Expanding beyond SARS-CoV-2 to other pathogens
7. Public Health Impact: Informs timely preventive strategies
8. Budget Estimate: Likely received tens of millions in funding from the broader $1.75 billion allocation

(2) 일본

여행객의 건강 상태를 기입한 「질문표」를 입국·귀국할 때에, 검역관에게 제출하도록 하고 있고 그 방법으로 일본에 도착하기 전에, 자택·출발지의 공항·항공기 내 등에서 「질문표 Web」에 의해 질문 항목을 입력해, QR코드를 작성해, 화면을 보존 또는 인쇄해 빠른 검역 수속을 할 수 있게 한다. 「질문표 Web」을 스마트폰이나 태블릿의 홈 화면에 추가하는 것으로, 항공기 내 등의 오프라인 환경에서도 「질문표 Web」의 입력이 가능하다. 또한 일본의 자발적인 검역 참여 프로그램의 하나인 게놈 감시는 5개 공항에서 유증상자를 대상으로 유증상자자 협조자 중에서 비강 검체채취를 실시하여 호흡기 감염증을 일으키는 주요 바이러스나 세균의 병원체 유전자를 전체적으로 검출할 수 있는 PCR 검사하여 게놈 해석 등을 실시한다.

(3) 한국

자발적인 신고서비스는 여행자가 건강상태질문서를 작성하여 제출하거나 Q-CODE를 활용하여 건강상태를 신고하고 있으며, '여행자 대상 호흡기감염병 검사서비스 시범사업'을 통해 해외감염병 및 신변종 병원체 유전체감시를 2025년 2월부터 시행 중에 있다.

4. 국제협력과 해외감염병 정보수집

1) 국제협력과 검역

WHO의 국제보건규칙(IHR)은 감염병의 국제적 확산을 방지하고, 공중보건의 위협을 효과적으로 대처하기 위해 제정되었다. 2005년 개정된 IHR은 196개국이 가입하여, 국가 간 협력을 통한 감염병 예방 및 대응 체계의 구축을 목표로 하고 있으며 검역과 매우 밀접한 관련이 있다. IHR의 주요 목적인 국경 간 감염병 확산 방지에 따라 감염병이 국제적인 공중보건 위기를 초래하지 않도록 조기 감지 및 차단하고 국가 간 정보 공유를 강조하고 있다. 각국은 감염병 확산 방지를 위해 국제적인 보건협력을 기반으로 감염병 발생과 확산에 대한 투명한 정보 제공과 실시간 보고 체계를 구축해야 한다는 것이다. 현대 사회는 글로벌화로 인해 국경을 초월한 이동이 활발하고 감염병은 항공, 해상 운송을 통해 빠르게 확산될 수 있다. 예를 들어, SARS와 SARS-CoV-2는 단기간에 세계 여러 지역으로 확산되며 전 지구적 공중보건 위기가 되었다. 감염병은 국경에서 차단하는 것이 핵심이지만, 한 국가의 노력만으로는 한계가 있고, 따라서 WHO IHR은 국가 간 긴밀한 협력을 통해 검역 시스템을 강화하고, 감염병 확산을 방지하는 데 중점을 두고 있다. 또한 IHR은 단순히 사람에 대한 검역뿐 아니라 화물, 운송수단, 동물 및 제품에 대한 관리도 포함하고 있으며 에볼라와 같은 감염병은 국경을 넘어 확산될 위험이 있으므로 검역소에서 철저한 관리와 통제를 강조하고 있다.

2) 여행자 기반 유전자 감시 프로그램

여행자는 이동성이 있고 여행 중에 노출될 가능성이 있으며 단시간 내에 한 장소에서 다른 장소로 질병을 퍼뜨릴 수 있으므로 새로운 감염병과 출현하는 감염병을 추적할 때 고려해야 한다. 해외 감염병 예방과 정보수집을 위한 양방향 정보시스템 구축을 구축해야 하고 대표적인 사례가 미국형 TGS CDC Traveler-based Genomic Surveillance Program (TGS)이다. 2021년 9월 출범 이래 Los Angeles, San Francisco, New York (JFK), Washington Dulles, Newark, Boston, Seattle, Miami, and Chicago에 설치되어 새로운 병원균을 조기에 감지하고 전 세계 생물 감시의 빈틈을 메우는 메커니즘으로 활용할 수 있는 역량을 보여주고 있다. 이와 같이 자발적인 입국 여행자 감염병 보고체계 구축으로 여행자에게 인센티브를 제공하는 시스템을 제도화한 미국 TGS는 해외 감염병의 전파 기전과 발생 양상을 과학적으로 평가하고 조기 경고와 대비를 할 수 있는 방역 전략을 수립하는 데 크게 기여할 것이다.

3 검역방법

1. 검역절차

1) 팬데믹: 공항검역

(1) 검역관리지역 검역

코로나19 팬데믹 감염병 유행 시 그림 7-1-5와 같이 모든 입국자를 대상으로 발열 체크, 건강상태 질문서 확인한 후 '특별입국 절차'를 수행하고 건강상태 질문서를 확인하는 과정에서 증상, 병원 방문, 약 복용 여부 및 검역감염병 감시 기간 내 방문·체류한 국가를 확인한다. 발열 확인은 발열 감시 카메라 및 비접촉 체온계 이용하고 입국하는 모든 내·외국인은 특별입국 절차에 따라 체온을 측정하고 기내에서 사전 배부받은 건강상태질문서와 특별검역 신고서를 작성해야 한다. 검역관은 국내 체류 주소 및 연락처를 확인하고, 입국자 본인 소지 휴대전화에 '자가격리자 안전보호 앱' 또는 '자가진단 앱' 설치 여부와 연락처를 현장에서 확인한다.

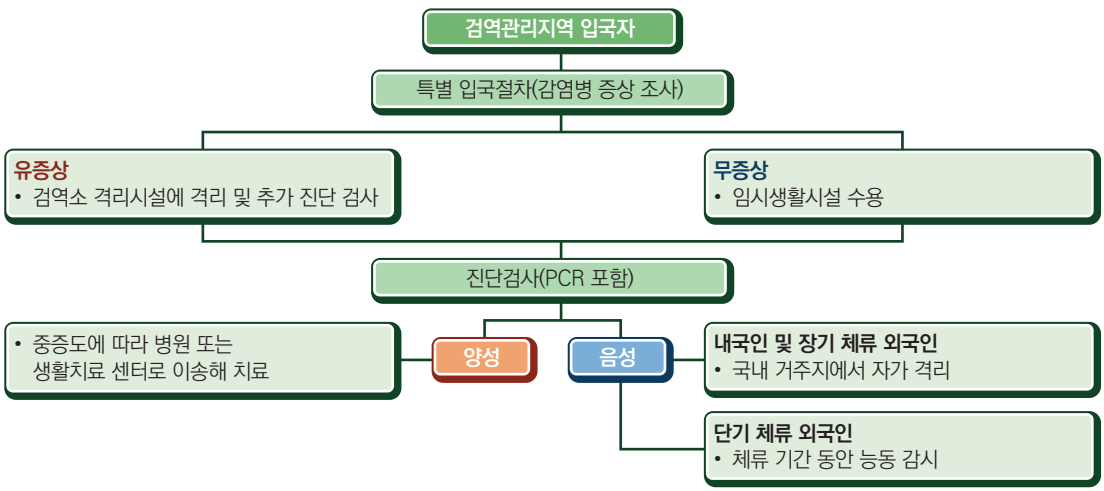


그림 7-1-5. 입국자의 검역절차(코로나19)

(2) 진단검사

해외에서 입국하는 모든 내·외국인은 입국 전·후 검사를 의무적으로 실시해야 한다.

⑶ 내국인·장기체류 외국인

자가격리 대상임을 안내하고 행정안전부는 '자가격리자 안전보호 앱' 설치를 확인 및 특별입국 절차 후 입국 조치를 안내한다.

⑷ 단기 체류 외국인

임시생활시설격리 대상임을 안내하고 '자가격리자 안전보호 앱' 설치 확인 후 (시설 입소 시 동 앱 삭제 후 '모바일 자가진단 앱' 설치) 입국 조치한다.

2) 팬데믹: 선박 검역

선박 검역의 서류검역과 승선검역 기준은 검역관리지역 지정 여부 외에 표 7-1-2와 같이 출항 국가의 위험도, 승선자의 접촉 강도, 국내 입국 하선자의 위험도에 따른 위험성 결정에 따른다.

표 7-1-2. 선박 검역의 위험도 평가지표 및 판단 기준

위험도 평가지표	판단 기준
출항국가의 위험도	코로나19 신규환자 발생 및 국내 유입 확진자 현황
입항 선박의 승선자 접촉 강도	대면 및 접촉이 필요한 하역, 수리 업무 수행 * 특히 비대면 하역이 가능한 선박인지 고려
국내 입국 하선자의 위험도	입항 전 검역관리지역에서 검역감염병의 최대 잠복기간 이내 선원 교대가 있었던 선박, 유증상자 발생 신고 선박 등

3) 평시 검역

코로나19 이후 검역체계는 여행객의 불편을 최소화하면서도 검역의 효과를 높이기 위한 방안으로 다양한 검역 수단을 개발하고 있다. 출입국 전단계에서 여행객에게 감염병 예방에 도움이 되는 해외 감염병 발생 동향이나 체류지의 건강 정보 서비스를 제공하고 자발적인 신고와 검사가 진행되도록 인센티브를 마련하는 등 국민 참여형 검역 체계로 개편이 이루어지고 있다. 평시 검역은 중점검역관리지역으로 지정한 곳에서 출발하는 여행객은 Q-CODE를 통하여 전수가 신고하여야 하고 그 밖의 지역은 자율적인 자진 신고체계로 운영한다. 유증상자는 반드시 검역조사를 받아야 하고 필요시 격리를 할 수 있지만 무증상자는 자발적으로 검사를 원하지 않은 경우는 귀가할 수 있다.

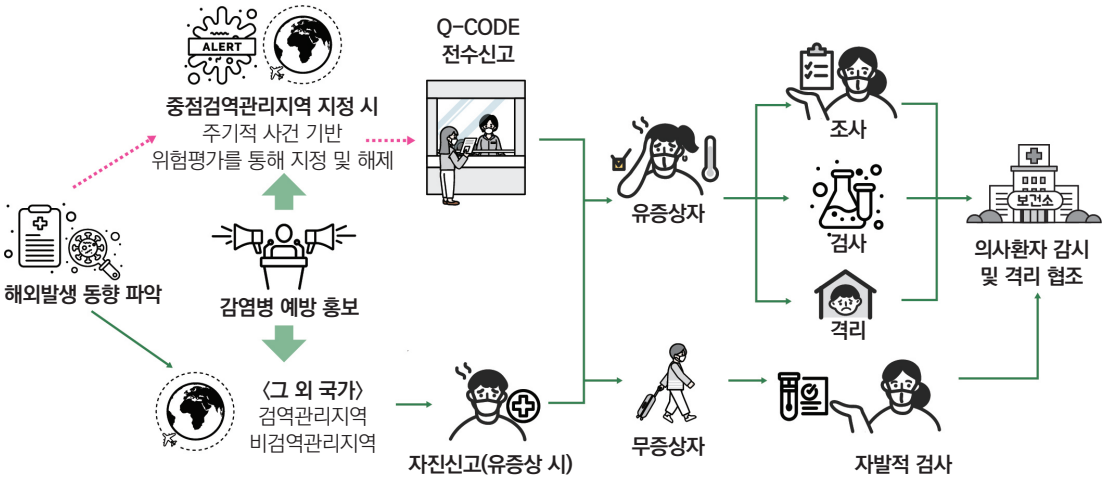


그림 7-1-6. 평시 출입자 검역체계

2. 국가별 검역체계

국가별 검역체계를 평시와 COVID-19 시기와 비교하면 표 7-1-3과 같다. 일본, 미국, 캐나다, 호주, 영국, 독일 등 많은 국가가 COVID-19 팬데믹 시기에 국경 봉쇄에 가까운 강력한 조치를 취하거나 엄격한 검역을 시행하는 것에서 평시에는 사전 검역 혹은 서류 검역 등 비대면 검역 등으로 여행자나 운송수단의 검역 절차를 간소화하고 있다.

표 7-1-3. 주요 국가별 검역체계 비교

국가	검역 절차 및 내용	평시	COVID-19 팬데믹 시기
일본	중점검역	• 법령규정 중점검역이라는 용어는 없지만 이에 해당하는 근거 조항 존재 • 황열 유행지역에 입국하는 경우 등에는 국제보건규정(IHR)에 따라 황열 예방접종증명서(옐로카드)를 제시 요구 • 검역법 제14조: 검역 감염증이 유행하고 있는 지역을 발항한 선박 및 항공기에 대해 전수 검역조치 존재	• 입국 거부 대상 국가 및 지역을 지정하여 운영:14일 이내에 한국 등 입국거부 대상 국가 및 지역에 체류이력이 있는 모든 입국자(일본인 포함) 대상 검역소장이 지정하는 장소에서 14일간 대기 및 대중교통 사용 자제 요청, 아울러 입국전 14일 이내 입국거부 대상지역에 체류 이력이 있는 입국자에 대해 PCR 검사 실시

국가	검역 절차 및 내용	평시	COVID-19 팬데믹 시기
일본	중점검역	• 검역법 제21조: 후생노동성령으로 지정하는 외국의 지역을 발항하거나 그 지역에 기항하여 내항한 경우 별도의 검역 절차가 있음	• 공중보건상의 관점에서 신종감염병 발생국·지역에서 내항하는 항공기·선박의 국내 검역실시 장소를 지정해 집약화를 도모하기 위해 검역집약공항 및 항만을 지정하여 대응(나리타, 간사이, 주부, 후쿠오카공항, 요코하마, 고베, 간몬항) • 전수조사 후 격리, 정류, 소독 수행
	발열 감시	실시	실시
	자발적 신고	실시 (질문표Web 전자화)	실시
	검역 관리지역	법령규정 없음	법령규정 없음
	오·하수 검사	검역법 제27조: 해당 지역 내에 있는 선박 청소년 또는 항공기에 대해 식품, 음료수, 오물, 오수, 쥐류 및 곤충류의 조사를 실시	실시
미국	중점검역	주요 공항 및 항만에서 주기적으로 실시	코로나19 확산국에서 오는 여행자 대상 집중 검역 시행
	발열 감시	없음	국제선 입국 시 발열 감시 의무화
	자발적 신고	여행자 건강 신고 시스템 활용	입국 전 건강 상태 신고 의무화
	검역 관리지역	지정된 국가에 대한 특별조치 없음	중국, 이란, 유럽 일부 국가 등 지정 후 입국 제한 및 검역 강화
	오·하수 검사	주요 공항 및 항만에서 제한적으로 시행	항공기, 선박에서 채취된 샘플에 대해 코로나19 검사 시행
캐나다	중점검역	미실시	PCR 검사 도착일로부터 14일간 의무 격리 규정을 준수하지 않는 여행객은 PHAC에 회부되어 추가 평가를 받고, 최대 3,000달러의 벌금과 형사 고발을 포함한 강제 조치를 받을 수 있음
	발열 감시	미실시	실시
	자발적 신고	실시	실시
캐나다	검역 관리지역	미실시	실시
	오·하수 검사	캐나다에서 공중보건 위험을 초래할 가능성이 있을 때 지방, 지역 및 지방 자치 단체에서 결정	캐나다에서 공중보건 위험을 초래할 가능성이 있을 때 지방, 지역 및 지방 자치 단체에서 결정
호주	중점검역	각 주와 준주에는 생물 보안 또는 격리 구역이 있음	COVID-19 동안 감염 환자는 호주 병원에 격리 검사를 받은 후 결과가 통보될 때까지(보통 48시간 이내) 격리

국가	검역 절차 및 내용	평시	COVID-19 팬데믹 시기
호주	발열 감시	(실시) 일부 항공사와 공항(Qantas & Virgin Australia)에서 검사 절차의 일환으로 열 스캐너 사용	(실시) 예방 조치로 열 스캐너를 사용하여 도착 승객을 대상으로 체온 검사를 시행
	자발적 신고	(실시) 동식물 및 식품을 반입할 때, 호주의 검역 규정에 따라 세밀한 검역을 시행함	실시
	검역 관리지역	각 주와 준주에는 생물보안 또는 격리 구역이 있음	실시
	오·하수 검사	(실시) 정기적으로 샘플링 진행	(실시) 폐수 감시를 통해 COVID-19를 유발하는 바이러스인 SARS-CoV-2를 모니터링
영국	중점검역	미실시	실시 입국과정에서 PCR 검사 개인 자가격리 준수 여부 전화로 확인
	발열 감시	미실시	실시
	자발적 신고	실시	실시
	검역 관리지역	미실시(IHR 30조 규정에 근거하여 여행자가 공중보건에 위험을 초래하지 않는다면 자유로이 여행을 계속할 수 있다는 근거로	실시(Managed Quarantine Service에서 Red list 운영)
	오·하수 검사	실시	실시
독일	중점검역	미실시	실시
	발열 감시	미실시	실시
	자발적 신고	실시	실시
	검역 관리지역	미실시	실시(RKI 홈페이지에 게시)
	오·하수 검사	실시	실시

3. 검역의 도전과 미래 전망

1) 감염병 해외유입 및 국내유행 위험 증가

전 세계가 하나의 생활권으로 변화함에 따라 검역환경도 항만에서 공항으로, 선박·물류에서 항공기·승객으로, 콜레라 등 세균성 감염병에서 메르스, 코로나19 등 바이러스 감염병으로 변화하고 있다. 그동안 해외에서 유입된 감염병인 사스(2003), 신종인플루엔자(H1N1, 2009), 메르스(2015), 코로나19(2019)가 국내 공중보건 위기 상황을 발생시켰다. 기후 환경의 변화로 인한 신종 감염병 및 재출

현 감염병의 위험이 증가하고 있고 20세기 이후 발생한 에볼라바이러스병(2017.5월), 크리미안콩고열(2016.8월), 서아프리카 라싸열(2016.2월) 등 신종 감염병의 대부분(75%)이 야생동물로부터 유래한 바이러스가 원인이 되고 있다. 더욱이 지구 온난화의 가속화로 모기, 진드기 등 감염병 매개체 서식지 확대로 인한 매개체 감염병[중증열성혈소판감소증(SFTS), 황열, 뎅기열, 지카바이러스 감염증 등] 증가 및 해외유입 위험이 증가하고 있다.

2) 검역조직혁신 및 정책 변화요구 증대

코로나19 이후 검역조사 체계 개선 요구 및 이를 위한 연구개발에 대한 수요가 증대되면서 질병관리청 내 권역별 질병대응센터 설립으로 검역소 역할 변화 요구가 증대되고 있고 감염병 대유행이 주기적으로 발생하고 있는 상황에서 감염병 유관기관의 조정·협력을 통한 지속 가능하고 실효성이 있는 대응시스템이 필요한 시점이다. 또한 매년 급증하는 입국자에 대비한 검역 절차에 대한 변화가 필요하다. 현행 중점검역관리지역을 최소화하여 핀셋 관리하고 검역관리지역은 정보제공 중심으로 자발적 신고기반검역으로 전환하여 여행자에게 편익을 주면서 검역의 효과를 높이는 검역행정이 되어야 한다. 검역의 방식에 있어서 해상검역이나 크루즈 검역을 상당부분 전자검역으로 대체하고 중점검역관리지역은 Q-CODE 제출을 의무화하는 등의 조치가 필요하다. WHO는 입국자가 여행지 노출을 회피하는 것을 방지하기 위해서 여행 후 의료정보 등 서비스제공 중심 검역으로 권장하고 있다.

표 7-1-4. 검역체계 개편 방향

구분	현행	개편방향
중점검역관리지역	운영	제한적·탄력적 지정 → Q-CODE 의무화 & 발열감시
검역관리지역	검역질문서 & 발열감시	정보제공과 신고기반 검사실시
비검역관리지역	발열감시	발열감시

참고 문헌

1. 질병관리청, 이윤현, 김명성, 남영희. 검역법령 개정방향 수립에 따른 세부규정 마련을 위한 연구. 2019.
2. 질병관리청, 이윤현, 남영희, 김명성, 정규석, 이윤석. 해외 주요국가 검역사례 및 법제도 연구. 2024.
3. 질병관리청, 이윤현, 남영희, 김명성. 국외감염병정보센터 구축에 따른 편익 효과 분석. 2023.
4. 질병관리청, 이윤현, 홍지영, 남영희, 김수연, 우세린, 신창섭. 검역관리 기본계획 및 검역조사 알고리즘 개발. 2020.
5. 질병관리청, 이윤현. 해외 주요국가 검역사례 및 법제도 연구, 질병관리청 정책연구 용역보고서. 2025.
6. AARP [Internet]. Traveling Overseas This Winter Consider? Testing for Public Health Monitoring. The CDC has ex-

panded its voluntary testing for respiratory viruses at airports. AARP. [cited 2024 Dec 26]. Available form: https://www.aarp.org/travel/travel-tips/transportation/info-2023/cdc-respiratory-virus-screening-at-airports.html.

7. CDC [Internet]. Traveler-based genomic surveillance for early detection of new SARS-CoV-2 variants Surveillance for Early Detection of New SARS-CoV-2 Variants. CDC Stacks. 2023. [Cited 2024 Dec 26]. Available form: https://www.aarp.org/travel/travel-tips/transportation/info-2023/cdc-respiratory-virus-screening-at-airports.html.
8. Center on Budget and Policy Priorities. Putting the Size of the Needed COVID-19 Fiscal Response in Perspective. 2020.
9. WHO [Internet]. Working Group on Amendments to the International Health Regulations(2005) [Cited 2024 Dec 5]. Available form: https://apps.who.int/gb/ebwha/pdf_files/WHA77/A77_ACONF14-en.pdf.

제2절 해외유입 감염병의 예방과 관리

1 해외여행과 감염병

감염성 질환은 숙주와 병원체를 둘러싼 환경 및 생태학적 특성에 크게 영향을 받는다. 역사적으로 감염성 질환은, 각 지역의 기후 및 생태학적 환경, 병원체의 병원소 및 숙주 역할을 하는 동물과 야생의 조건, 숙주의 생활습관 및 건강행태 그리고 이들의 상호작용에 따라 지역적으로 다양한 양상으로 변화해 왔다. 이러한 결과는 지구상의 각 지역마다 그 지역의 생태학적, 문화적 특성에 따라 다양한 풍토병(endemic diseases)을 유행시키게 되었다. 예를 들면 중부 아프리카의 뇌수막염 벨트, 동남아시아의 말라리아 유행지역, 남아메리카의 황열 유행지역 등등 각 지역마다 고유의 풍토병들이 있다. 과거에는 이러한 각 지역의 풍토병은 그 지역의 문제이거나 그 지역을 여행하는 소수의 문제였다. 그러나 현재는 인적, 물적 국제교류가 매우 커지고 빈번해지면서 지구 전체가 하나의 생활 영역으로 묶이고 지구촌 어느 지역의 풍토성 감염병도 이제 더 이상 강 건너 불이 아닌 상황이 되었다. 과거 동남아시아와 아프리카 일부의 풍토병이었던 뎅기열이 현재 위도 0~±30도인 지역의 대다수 국가에 유행하는 감염병이 되었고, 중부아프리카 특정 지역에서 발견된 지카바이러스가 아프리카를 넘어 서태평양지역, 남아메리카, 중부아메리카로 확산되었다.

UN과 국제여행기구(World Travel Organization)에 의하면 전 세계 국제여행객 수는 2005년에 8억 8백만 명에서 2010년에는 9억 1백만 명, 2015년 11억 5천만 명 그리고 코로나 유행 전 2019년에는 14억 2천만 명으로 가파르게 증가하였다(그림 7-2-1). 코로나 유행기인 2020년과 2021년은 각각 4억 6백만 명, 4억 6천만 명 수준으로 크게 감소하였다가 이후 다시 빠르게 회복하여 2024년에는 2019년 수준인 14억 5천만 명으로 회복하였다. 앞으로도 이 수치는 계속 증가할 것으로 예상된다.

우리나라 국민의 국제여행객 수도 2005년 9백만 명 수준에서 2008년 1,332만 명으로 증가 후 2009년 신종플루 등의 영향으로 2010년 950만 명 수준까지 감소하다가 이후 계속 증가하여 2015년 1,931만 명, 2019년 2,871만 명으로 증가하였다(그림 7-2-2). 코로나 시기인 2020년 428만 명, 2021년에는 122만 명으로 2019년의 5% 정도까지 떨어지기도 하였으나, 2024년에는 2,869만 명으

로 다시 유행 이전으로 빠르게 회복하였다. 인구수 대비 단순하게 비교하면 국민 2명당 1명 꼴로 해외여행을 하고 있는 셈이다. 또한 우리나라를 방문하는 해외여행객 수도 2015년 1,323만 명에서 2019년 1,750만 명으로 증가했다가 코로나 유행기인 2020년 97만 명 수준으로 감소하기도 하였으나 2024년 1,103만 명 수준으로 증가 추세이다. 우리나라의 해외여행객 수도 지속적으로 증가할 것으로 예상되며, 여행하는 지역도 더욱 다양해질 것이다. 비단 사람뿐 아니라 각종 동식물을 포함한 전(全) 지구적인 물류 교류의 증가는 감염병에 있어서도 세계화를 가속화할 것이기 때문에 해외감염병 관리는 계속 중요성이 커질 것으로 판단된다.

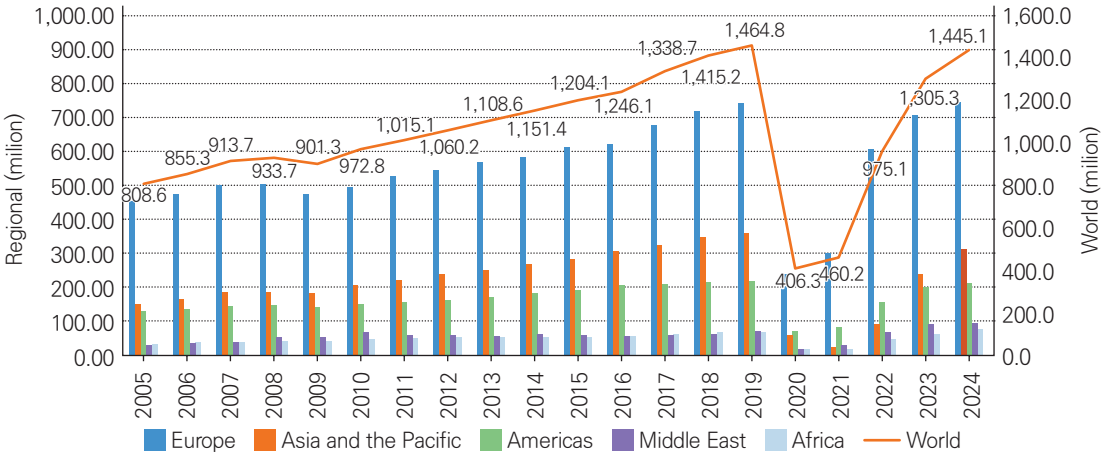


그림 7-2-1. 2005~2024년 전 세계 국제여행객 수(도착기준)
[출처: UN tourism/Statitiscs]

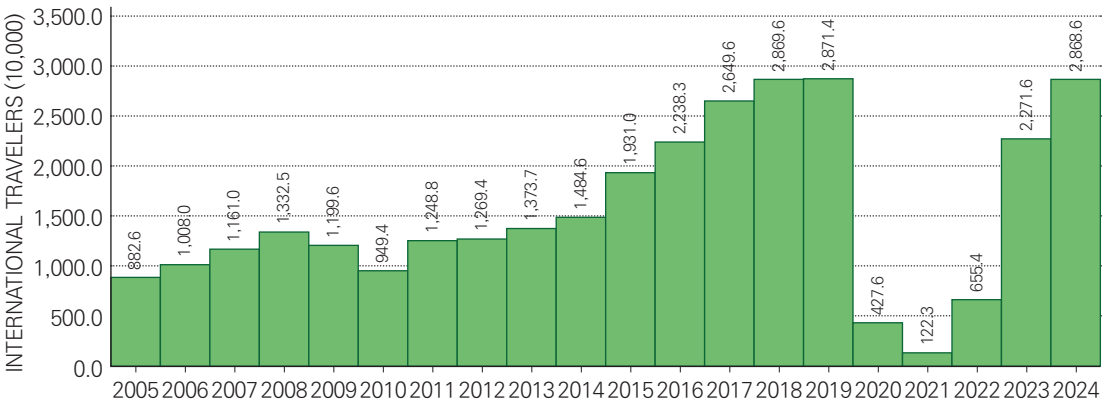


그림 7-2-2. 2005~2024년 한국의 국제여행객 수
[출처: 한국관광데이터랩, https://datalab.visitkorea.or.kr/datalab/portal/main/getMainForm.do]

2 해외감염병 감시체계

해외유입 감염병을 예방하고 관리하기 위해서는 적어도 두 가지 정보가 필요한데, 첫째는 어느 지역에서 어떤 질환이 유행하고 있는가에 대한 정보이고 두 번째는 그 지역에 여행 또는 거주 시 해당 감염병에 걸릴 위험이 얼마나 되는가이다. 이 두 가지 정보는 각 감염병의 유입위험도 평가나 우리나라 국민이 여행시 어떤 질환을 얼마나 주의해야 하는지 교육하는 데도 필수적인 정보라고 할 수 있다. 예를 들어, 같은 풍토병이라고 하더라도 해당 주민에서도 발생수준이 매우 낮은 것과 외부인이 1주일 여행 시 10%의 감염위험이 있는 질환은 그 위험도가 매우 다르다.

같은 국가라고 하더라도 지역에 따라 유행하는 감염병은 다른 경우가 많아서 구체적인 지역별 발생상황을 아는 것도 중요하다. 또한 지역별로 알아야 하는 감염병은 해당 지역의 고유한 풍토병뿐 아니라 세균성이질이나 홍역처럼 거의 모든 국가에서 발생하는 질병이 현재 유행하고 있는지에 대한 정보도 필요하기에 실시간 정보가 필요하다고 할 수 있다. 이러한 정보를 수집하는 체계는 각 국에서 보고하거나 매스컴을 통해서 알려지는 경우, 전문가 네트워크를 통해서 알려지는 경우 등도 있고, 해당 지역을 여행하는 사람들을 조사하는 모니터링 체계를 가진 경우도 있다. 이러한 감시체계들을 크게 해외감염병 발생정보와 발생 위험도 평가정보로 나눌 수 있다. 또한 해외감염병 감시체계도 자료의 특성에 따라 1차자료를 이용한 감시와 2차자료를 이용한 감시로 나눌 수 있다. 1차자료의 경우는 감시를 목적으로 여행자 질병감시체계를 운영하는 것이고, 2차 자료는 매스컴이나 뉴스 등의 발생자료를 수집하는 체계이다. 1차 자료를 이용한 해외감염병 감시는 GeoSentinel이 거의 유일하다.

1. GEO-Sentinel 네트워크

Geo-Sentinel은 해외 여행자와 이주민을 대상으로 감염병을 감시하는 글로벌 네트워크로, 국제 보건기구 및 연구자들에게 중요한 역학 정보를 제공한다. 1995년부터 시작된 이 네트워크는 국제여행의학회(International Society of Travel Medicine, ISTM)와 미국 CDC (Centers for Disease Control and Prevention)에서 공동으로 운영하고 있는데 전 세계 29개국에 있는 71개의 여행클리닉, 감염병 전문병원, 대학병원 등에서 이 감시 네트워크에 참여하고 있다(그림 7-2-3). GeoSentinel은 각 참여기관을 이용하는 해외여행자들의 여행관련 질환과 여행력을 보고함으로써 언제 어디를 여행한 여행자에서 어떤 질환이 발생했는지를 알 수 있게 해준다. 이러한 감시는 단순히 어디에서 어떤 질환이 발생하는가뿐 아니라 여행을 했을 때 해당 질병에 걸릴 위험이 얼마나 되는지도 평가할 수 있는 자료를 제

공한다. GeoSentinel을 통해서 얻어진 표준화된 해외여행 관련 질환의 정의와 수집정보는 해외여행 관련 질병역학연구에 많은 기여를 하고 있는데, 지역별, 계절별, 질병별, 해외여행의 목적과 여행자의 특성별 구체적인 질병위험에 대한 많은 자료와 논문을 산출하고 있다.

예를 들면, 유럽에서 2008년에서 12년까지 수집된 해외여행감염병 32,136명을 분석하여 여행 목적이나 여행자 특성에 따라서 여행관련 감염병의 상대유병률과 말라리아예방약의 효과를 평가하여 제시하고 있는 것을 볼 수 있다. 또 GeoSentinel에 등록된 17,353명의 분석에서는 전 세계 여행지를 캐리브해 지역, 중앙아메리카, 남아메리카, 사하라남부 아프리카, 남중아시아, 동남아시아, 기타 지역으로 나누어서 각 지역을 여행 후 여행의학 클리닉을 찾은 환자의 다빈도 증후군과 그 증후군의 다빈도 원인이 무엇인지를 분석해서 보고하였다. 이러한 정보는 단순하게 어떤 지역에 어떤 질환이 유행하고 있다는 매스컴이나 정부의 보고와는 다른 중요한 정보를 제공한다. GeoSentinel의 지역조직으로 유럽에서 참여하는 여행의학 클리닉들은 따로 EuroTravNet을 운영하고 미국의 경우는 GTEN (Global TravEpiNet)을 운영하고 있다. 우리나라는 2025년 2월 현재 1곳의 병원이 참여하고 있는데 한국인의 해외여행질환 특성을 파악하기에는 부족해 보인다.

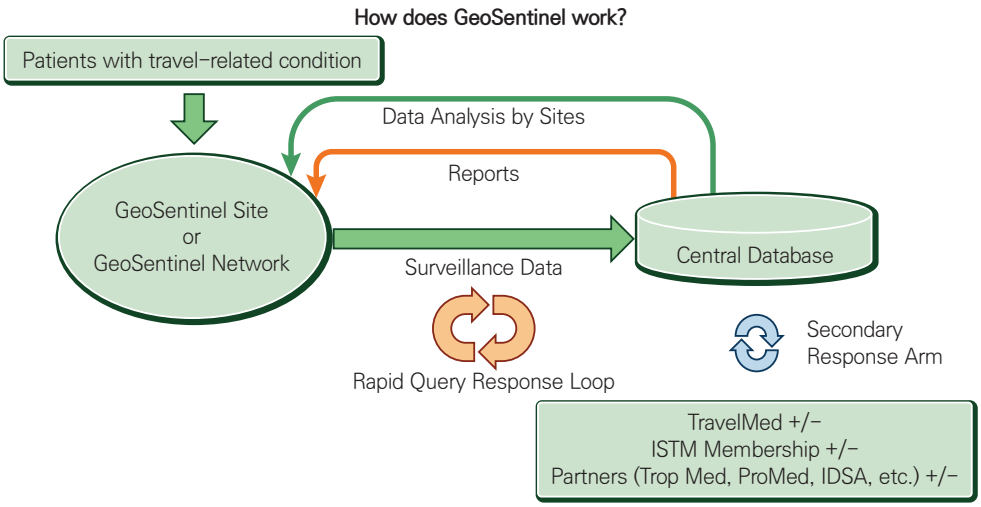


그림 7-2-3. GeoSentinel 네트워크의 감시체계 구조
[출처: https://geosentinel.org/about/structure/#single_column_content-12]

GeoSentinel과 유사하게 스위스내 대학에서는 WHO의 여행의학 협력센터로서 1984년부터 각종 자료를 종합하여 해외여행관련 위험도를 평가하고 있다. 예를 들면 개발도상국가를 1달 여행시 여행자 설사는 20~60%, 말라리아(예방화학요법 없이 서부아프리카 방문)는 2~3%, 뎅기열은 1%, 광견병 위

험이 있는 동물에게 교상 입을 확률 0.4~0.5%, 결핵 0.4% 등등 각 질병별로 위험도를 평가하고 지속 업데이트하여 보고하고 있다. 이 자료는 해외여행관련 감염병의 발생률을 측정하고 있다는 점에서 귀중한 자료이다. 다만 이 자료들은 유럽인이나 미국인들 해외여행객 중심으로 진행되는 감시자료로 우리나라 국민에 적용하기에는 한계점이 있을 것으로 사료된다.

2. ProMED 네트워크

ProMED (Program for Monitoring Emerging Diseases)는 국제감염학회(International Society for Infectious Diseases, ISID)에서 운영하는 전문가 네트워크이며 2차 자료를 활용한 감시체계이다. 1994년부터 시작해서 30년이상 하루도 빠짐없이 활동해 오고 있는 이 네트워크는 전 세계 어디에서 감염병이나 원인 모를 질환의 유행이 공식적인 매스컴에 나오거나 보고가 된다면 이 ProMED 네트워크를 통해서 전 세계의 감염 관련 전문가에게 이메일로 전달된다. 여기에는 사람 감염병뿐 아니라 동물의 감염병 유행도 포함되며, 어떤 네트워크보다도 유행을 신속하게 전달하고 있으며, 거의 전 세계 유행을 다루면서 하루에도 20~30개씩의 유행이 보고가 되고 있어서 전달하는 양도 방대하다. 즉, 특정 지역에 현재 어떤 질환이 유행하고 있는지를 알고자 하면 이 ProMED 네트워크의 자료를 보면 된다. ProMED에서 제공하는 유행지도에서 검색도 가능하다(그림 7-2-4).

ProMED와 연계된 전문가 네트워크로 EpiCore가 있다. EpiCore 프로그램은 ProMED, TEPHINET, HealthMap 등의 프로그램과 연계해서 각 지역의 전문가들을 선정해서 해당 지역에 원인 모를 질환이 유행한다는 소식이 있으면 그 지역의 EpiCore 전문가에게 즉시 확인을 요청하고 그 결과를 전 세계 전문가들에게 알린다.

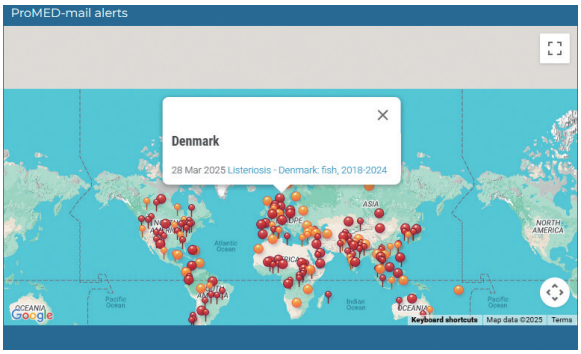


그림 7-2-4. ProMED 홈페이지에서 제공하고 있는 감염병 유행지도
각 핀을 누르면 구체적인 정보와 관련 메일의 링크가 뜬다.

3. HealthMap 및 기타

HealthMap도 2차 자료를 활용한 감시체계로 미국 하버드대학교 보스턴 소아병원의 계량역학연구실(Compuational Epidemiology Lab)에서 운영하는 다양한 감염병지도 프로젝트 중 하나이다. 감염병 유행과 관련된 각종 2차 자료를 전세계에서 수집하고 이를 mapping하여 이용자 중심의 자료를 제시해준다. 제공하는 정보도 ProMED만큼 방대하여 통상 1주에 1,000건이 넘는 정보를 제공하고 있다. HealthMap은 이용자 주위의 감염병에 대한 경보, 이용자 주위에 백신을 접종할 수 있는 기관 안내, 인플루엔자 지도, 뎅기열 지도 등 다양한 정보를 수집하고 이를 지도에 표시하여 이용자가 어느 위치에 있든 주위의 정보를 제공하고 있다.

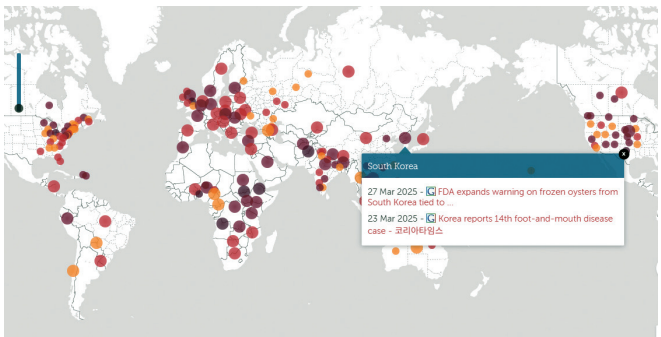


그림 7-2-5. HealthMap 제공 감염병 유행발생 지도

이외에 GIDEON (Global Infectious Diseases and Epidemiology Network)은 세계 각국의 감염병 유행 및 관련 데이터베이스를 수집하여 유료로 제공하고 있다.

그러나 이러한 감시체계는 관련기관이나 전문가를 위한 것으로 볼 수 있으며 영어로 제공되어 있어서 일반 한국인 해외여행자가 이용하는 것은 쉽지 않다.

3 우리나라 해외유입 감염병 현황

2024년 현재 우리나라의 경우, 법정감염병 신고체계에서 신고된 질환에 대해서 국내 감염과 국외 감염으로 구분하여 해외유입 감염병에 대한 통계를 작성하고 있다. 또한 샤가스병, 광동주혈선충증, 유극악

구충증, 사상충증, 포충증 등에 대해서는 각 의과대학의 기생충학교실(또는 열대의학교실)을 중심으로 표본감시체계를 구성하여 감시하고 있다. 그렇지만 이러한 자료들은 해외유입 감염병의 위험도를 평가할 수 있는 분모에 대한 정보는 없다. 즉, 어느 지역에 어느 정도 머무는 사람이 몇 명인데 이 중에서 몇 명이 발생했는지는 알 수가 없다. 또한 법정감염병 신고자료이다 보니 해외여행 감염증의 가장 많은 비중을 차지하는 여행자설사, 피부질환, 인플루엔자 등의 경우는 자료가 없다.

표 7-2-1. 2018~2023년 해외유입 감염병 신고현황*

해외유입 감염병	합계	2018	2019	2020	2021	2022	2023
신종감염병증후군	17,337	0	0	5,379	11,958	–	–
코로나 19	62,588	–	–	–	–	55,855	6,733
뎅기열	786	159	273	42	3	103	206
말라리아	310	75	74	29	20	38	74
세균성이질	289	145	106	9	0	2	27
장티푸스	160	92	44	5	0	2	17
홍역	104	5	86	6	0	0	7
C형간염	66	28	17	7	5	4	5
치쿤구니아야열	41	3	16	1	0	8	13
장출혈성대장균감염증	40	15	18	2	0	1	4
라임병	28	4	11	1	0	5	7
유비저	17	2	8	1	2	2	2
지카바이러스감염증	11	3	3	0	0	3	2
Mpox	10	–	–	–	–	3	7
백일해	7	1	1	0	0	0	5
콜레라	3	2	1	0	0	0	0
기타	212	63	94	13	1	6	35
합계(모두)	82,009	597	752	5,495	11,989	56,032	7,144
합계(신종감염병증후군 및 코로나19제외)	2,084	597	752	116	31	177	411

* 연도나 출처마다 조금씩 숫자가 다른 경우 가장 최근의 자료를 기준으로 정리함.
출처: 질병관리청 백서(2023년, 2022~2021년 합본), 감염병포털

표 7-2-1에서 보면 2020년과 2021년 신종감염병증후군으로 신고된 것은 코로나19가 발생하면서 분류된 것으로 보이고, 코로나19 유행 시 이에 대한 검역의 강화로 코로나19가 해외유입 감염병으로 가장 많다. 이것을 통해서, 팬데믹 시기에 해외여행의 감소에도 상당수의 호흡기감염병 감염자가 해외로부터 유입되었는 지를 간접적으로 알 수 있다. 또한, 신종감염병증후군이나 코로나19를 제외하면

2020~2023년간 신고된 감염자 수가 크게 감소하는데 이는 해외여행객이 급감하면서 나타난 것으로 이후 여행객 수가 급증함에 따라 다시 가파르게 상승할 것으로 판단된다.

코로나19를 제외하면 이 기간동안 뎅기열이 786명으로 가장 많고, 다음이 말라리아(310명)로 모기매개감염병이 수 위를 차지하고 있는 것을 알 수 있다. 이와 함께 치쿤구니야열 (41명), 지카바이러스 감염증(11명) 등의 모기매개감염병도 꾸준히 신고되고 있다. 모기매개감염병 다음으로는 세균성이질(289명), 장티푸스(160명), 장출혈성대장균감염증(40명), 콜레라(3명) 등 식품매개감염병이 두 번째 흔한 해외유입 감염병으로 신고되고 있다. 이외에 홍역(104명), 백일해(7명) 등 호흡기감염병, C형간염(66명), Mpox(10명)와 같은 사람 간 접촉에 의한 감염병, 라임병(28명)과 같은 진드기매개감염병도 지속 신고되고 있다.

이들 해외유입 감염병의 유입지역은 아래 표 7-2-2와 같다. 코로나19 유행 전인 2018년, 2019년은 아시아 지역이 85% 이상을 차지했으나 코로나 유행기간인 2020~2022년은 33~54%정도의 비율이었고, 2023년도에 76.4% 정도로 증가했다. 이는 호흡기감염병의 경우 아시아지역뿐 아니라 다른 지역에서도 많은 수가 유입될 수 있음을 시사한다. 또한 이 수치는 지역별 여행객 수에 영향을 받는 것으로 아시아외 여행객 수가 많아지면 지역별 비중이 조금씩 달라질 것으로 판단된다.

표 7-2-2. 2018~2023년 해외유입 감염병 신고환자의 유입지역

	2018	2019	2020	2021	2022	2023
합계	597	752	5,495	11,989	56,037	7,122
아시아	520	647	1,835	5,724	30,130	5,444
아메리카	7	10	1,653	2,527	10,736	499
유럽	14	19	1,432	2,210	11,303	842
기타	56	76	575	1,528	3,868	337

출처: 질병관리청 백서(2023년, 2022~2021년 합본)

표 7-2-3은 2010년~2024년 해외유입 기생충질환 표본감시 결과인데, 이 기간동안 신고된 248명 중 톡소포자충증이 231명으로 93.1%를 차지하고 있다. 이 외에 포충증(7명), 주혈흡충증(4명), 리슈만편모충증이 간헐적으로 신고되고 있다. 이 질환들은 진단이 쉽지 않기 때문에 숨겨진 감염자가 있을 것으로 생각된다.

표 7-2-3. 2010~2024년 해외유입 기생충질환 표본감시 결과

구분	합계	'10	'11	'12	'13	'14	'15	'16	'17	'18	'19	'20	'21	'22	'23	'24
계	248	1	3	2	2	2	8	16	9	21	23	15	39	27	40	40
톡소포자충증	231	0	2	1	2	2	6	15	8	17	21	15	38	27	39	38
포충증	7	1	1	0	0	0	0	0	0	3	0	0	0	0	1	1
주혈흡충증	4	0	0	1	0	0	0	1	1	0	0	0	0	0	0	1
바베스열원충증	3	0	0	0	0	0	1	0	0	0	2	0	0	0	0	0
리슈만편모충증	2	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
사상충증	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
아프리카수면병	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
샤가스병	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
광동주혈선충증	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
악구충증	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
메디나충증	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

출처: 감염병 포털

4 해외유입 감염병의 예방

해외유입 감염병을 예방하기 위한 첫 단계는 국내 여행자가 해외여행 전에 목적지의 감염병 관련 정보를 충분히 알고 예방조치를 취하는 것이다. 이 예방조치들은 여행 전 상담, 예방접종 및 항말라리아제 등의 예방화학요법, 여행 후 상담 등을 포함한다. 이러한 예방조치들은 여행의학 클리닉에서 진행하는데 여행의학 클리닉을 통해서 백신과 예방약으로 예방할 수 있는 감염성질환이 많고, 여행 후 감염병의 조기 발견을 통한 유행 확산 방지 및 적절한 치료를 제공할 수 있기에 해외유입 감염병 관리에 중요하다. 그러나 50% 이상의 여행자들이 여행 전 여행의학 클리닉을 방문하고 있는 유럽과는 달리 아직 우리나라 해외여행자의 상당수는 황열접종 등 몇 가지 목적 외에 여행 전 여행의학 클리닉을 방문하지 않고 있기에, 정보 제공이 더 중요하다.

여행지역별 주의해야 할 감염병과 예방조치에 대한 정보들이 잘 제시되어 있는 곳은 다음과 같다.

1. 질병관리청 해외감염병 NOW

우리나라 질병관리청에서 운영하며, 해외감염병 예방을 위한 국가별 정보와 다양한 건강정보를 제공하고 있다. 한글로 제공되기 때문에 이용에 편리하다. 그러나 국가별 각 지역의 세부정보나 실시간 유행정보 등은 다른 정보를 같이 활용하도록 한다.

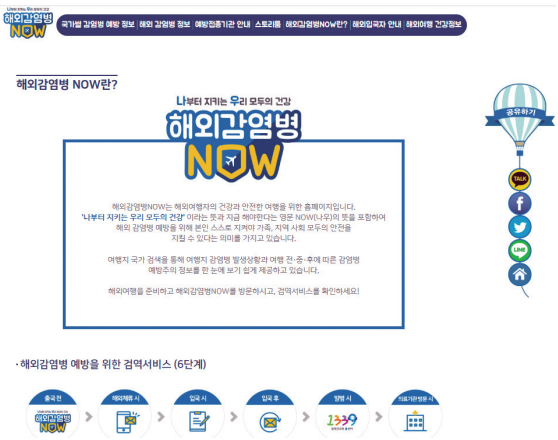


그림 7-2-6. 질병관리청 해외감염병 NOW 소개 페이지

2. 미국 CDC 해외여행자 건강 페이지

미국 CDC의 여행자 건강 페이지에서도 국가별로 다양한 감염병 관련 정보와 예방정보를 제공한다. 한 국가라도 모든 지역의 감염병 발생위험이 같지는 않기 때문에 구체적인 지도와 해당 지역을 제공하며 예방에 필요한 감염병별 예방조치와 주의사항에 대한 상세한 정보도 제공된다. 다른 CDC 정보처럼 일반인을 위한 정보와 임상가나 전문가를 위한 정보로 구분하여 제공하고 지역마다 실시간 유행정보도 제공하고 있다. 이러한 정보는 GeoSentinel 네트워크를 운영하면서 직접 얻기도 하고 다양한 자료원을 활용하고 있는 것으로 보인다. 또한 CDC Yellow Book이라는 매우 상세한 여행관련질환 책자를 제공하는데 정기적으로 업데이트되는 이 책은 여행의학관련 다양한 서적들 중에서도 가장 체계적이고 상세한 정보를 담고 있다.

CDC의 여행자 건강은 감염병에 국한되지 않고, 고산병이나 지역의 지형 특성, 문화적 차이, 운송수단에 따른 건강위험도 등도 같이 다루기 때문에 여행자 입장에서 더욱 유용하다.

3. WHO의 여행건강 페이지

WHO의 여행건강 페이지에서는 여행 전, 여행 중, 여행 후 관리에 대한 일반적인 원칙과 함께 DONs (Disease Outbreak News)를 통해서 실시간 유행정보를 같이 제공한다. 2012년 WHO는 소위 Greenbook이라고 하는 "International Travel and Health"라는 해외여행자 건강관련 좋은 책자를 낸 바 있는데 미국 CDC의 Yellowbook처럼 자주 업데이트가 되지 못하였었다. 그러나 2024년 7월에 말라리아, 예방접종대상 감염병, 환경적 위험, 신경학적 정신적 위험 등 주요 여행관련 질환별 모듈을 펴냈다. 다만, 해외여행의 경우 여행지역에 따라서 다양한 위험도 평가와 예방지침을 내리기 때문에 이처럼 질병별로 구분된 형태는 활용에 제한적인 부분이 있다.

이 외에도 세계 각 국가의 여행자들의 건강을 지키고 감염병을 예방하기 위해 활용할 수 있는 정보는 세계의 다양한 기관에서 제공하고 있다. 그러나, 우리나라 국민이 쉽게 접할 수 있는 국문 정보는 질병관리청 해외감염병 NOW 등으로 제한적이다.

4 해외유입 감염병의 관리

우리나라의 경우 해외유입 감염병의 관리는 검역과 입국자 추적 관리 시스템으로 수행하고 있다. 입국자 추적 관리 시스템은 입국자가 추적대상자에 해당하는 경우 지역사회와 연계하여 관리하는 체계이다. 추적대상자를 정리하면 아래와 같고 추적조사는 각 질병의 관리지침에 따른다.

- 입국 시 승객·승무원에서 콜레라, 장티푸스, 파라티푸스, 세균성이질, 장출혈성대장균감염증 병원체가 발견(확인)되거나 2인 이상 집단 설사 환자가 발견(확인)된 경우, 운송수단의 오수 등 가검물에서 콜레라균이 검출된 경우
- 입국 시 모기매개감염병 유증상자(말라리아 및 뎅기열 치료이력 자진신고자 포함)가 발견(확인)되는 경우
- 「제1급감염병 중동호흡기증후군(MERS)·중증급성호흡기증후군(SARS) 대응지침」, 「해외감염병 검역대응 표준매뉴얼」, 「제1급감염병 바이러스성출혈열 대응지침」 등에 따른 환자, 의사환자가 발견(확인)된 경우
- 에볼라바이러스병 위험지역에서 최대 잠복기(21일) 이내에 입국한 경우

또한 입국 당시 증상이 없었으나, 귀가 후 증상이 발현되어 국내에서 추적조사 대상 감염병의 환자(의심)로 확인된 경우는 역추적 조사대상이 된다.

이들 각각의 추적관리의 주체와 방법에 대해서는 해당 지침을 참조한다.

5 맺음말

해외유입 감염병은 코로나19 시기 여행자 수의 급감으로 같이 감소했으나 이제 빠르게 여행자 수가 증가함에 따라 다시 증가할 것으로 충분히 예상할 수 있다. 현재 우리나라의 해외유입 감염병 감시체계는 법정감염병 신고자 중 해외발생자나 해외유입기생충 질환에 대한 표본감시를 들 수 있다. 그러나 이들 감시체계에서 나오는 정보는 해외유입 감염병 예방과 관리에 활용하기에는 제한적인 측면이 있다. 우리나라 해외여행객이 여행하는 곳이나 국내 유입 외국인의 지역에서 어떤 감염병이 유행하고 있는지, 그 위험(risk)의 크기는 얼마인지 등을 좀 더 능동적으로 모니터링 해야 한다. 유럽 국가들의 경우는 오래 전부터 해외여행 감염병의 감시체계를 통해서 유럽인들이 언제 어느 지역을 방문할 때 여행을 얼마나 오래하는지 그 기간에 따라서 어떤 질병이 얼마나 걸릴지를 추산해 왔다. 즉, 해외여행 지역별로 여행자가 감염병에 걸릴 위험도(risk)를 추산하고, 그에 따른 발생률(incidence rate)을 계산해온 것이다. 우리나라의 경우 아직 이러한 감시체계를 구축할 수 있는 여행의학 클리닉이 많이 활성화되어 있지는 않아서 이러한 감시체계를 꾸리는 것이 현실적으로 어려운 것이 사실이다. 그러나, 그만큼 여행자들에게 더 정확하고 많은 정보를 제공해 줄 수 있어야 한다.

여행지별로 여행시기마다 감염병 발생 위험이 다르고, 감염위험은 여행 목적이나 여행자의 건강행태에 따라서도 상당히 달라진다. 지역마다 여행시기에 따라서 어떤 감염병 위험(risk)이 얼마만큼 있는가에 대해서 측정하고 어떤 위험요인이 해외에서 감염의 위험을 높이는지 등을 연구하는 분야가 여행질환 역학(travel epidemiology)이라고 할 수 있는데 아직 우리나라의 경우는 이 분야가 활성화되어 있지 않다. 국민의 건강을 담당하는 역학조사관으로서 이 분야에 관심을 갖고 발전시키는 것이 필요하다.

참고 문헌

1. 천병철. 해외여행 관련 질환의 역학과 여행의학의 감염병 관리. 대한의사협회지. 2010;53:510-23.
2. 해외감염병 NOW [인터넷]. 질병관리청. 2025 [Cited 2025 April 2]. Available from: http://해외감염병now.kr/nqs/oid-

now/main/index.do.

3. EpiCore [Internet]. HOW IT WORKS [Cited 2025 April 2]. Available from: https://epicore.org/how.
4. Freedman DO, Weld LH, Kozarsky PE, et al. Spectrum of disease and relation to place of exposure among ill returned travelers. N Engl J Med. 2006;354(2):119-30.
5. Gautret P, Schlagenhauf P, Gaudart J, et al. Multicenter EuroTravNet/GeoSentinel study of travel-related infectious diseases in Europe. Emerg Infect Dis. 2009;15(11):1783-90.
6. GIDEON [Internet]. GIDEON [Cited 2025 April 2]. Available from: https://www.gideononline.com.
7. HealthMap [Internet]. HealthMap [Cited 2025 April 2]. Available from: https://www.healthmap.org/en/.
8. ISTM [Internet]. Research&Publications [Cited 2025 April 2]. Available from: https://www.istm.org/geosentinel.
9. ProMED [Internet]. ProMED [Cited 2025 April 2]. Available from: https://promedmail.org/.
10. Steffen R, Amitrigala I, Margot M. Health risks among travelers-need for regular updates. J Travel Med. 2008;15:145-6.
11. Steffen R, Chen LH, Leggat PA. Travel vaccines-priorities determined by incidence and impact. J Travel Med. 2023;30(7):taad085.
12. Steffen R, Hamer DH, Chen LH, et al. Novel chikungunya and dengue vaccines: travel medicine applications. J Travel Med. 2024;31(4):taae064.
13. Travelers' health [Internet]. Centers for Disease Control and Prevention [Cited 2025 April 2]. Available from: https://wwwnc.cdc.gov/travel/.

역학조사관 교육
표준교재

Textbook For
Korea Field Epidemiology
Training Program





8장

원헬스와 기후변화







제1절 원헬스개념과 구성요소
제2절 인수공통감염병의 원헬스적 접근
제3절 항생제내성 관리와 원헬스
제4절 기후변화와 감염병

집필자 김홍빈, 유대성, 유석현, 정해관, 천병철
검토자 김종희, 신나리, 이선영, 이소담, 이신영

제1절 원헬스개념과 구성요소

1 원헬스개념과 구성요소

원헬스(onehealth)는 사람, 동물, 환경이 서로 밀접하게 연결되어 있고, 이러한 다양한 분야의 통합적이고 다학제적인 접근을 통해서 감염병의 예방 및 통제, 생태계 보호, 인류와 동물 건강 증진을 목표하는 접근방식을 말한다.

1. 원헬스의 개념

원헬스적 접근법은 감염병의 예방 및 관리, 식품위생, 항생제내성문제, 그리고 기후변화로 인한 건강문제와 같은 복잡하고 상호 연관된 건강문제를 주로 대상으로 한다. 특히, 2014~2015년 서아프리카 3개국에서 대규모 발생이 있었던 에볼라바이러스병, 2015년에 국내에 유입되었던 중동호흡기증후군, 2016년 브라질 등 남미지역에서 소두증의 원인으로 지목된 지카바이러스, 중국 및 동남아에서만 제한적으로 보고되다가 점차 그 발생이 확대되고 있는 조류인플루엔자 인체감염증 등 과 같은 신종감염병의 발생은 야생동물 서식지의 파괴, 인간과 동물의 접촉 증가 등으로 인해 영향을 받은 것으로 볼 수 있다. 따라서, 이러한 신종감염병의 경우 단일 기관 또는 단일 분야의 노력으로는 해결하기 어려운 문제이며, 이를 체계적이며 효과적으로 다룰 수 있도록, 다학제적 협력을 강조하는 실천적 개념으로 원헬스적 접근을 통한 감염관리 및 예방전략 수립의 필요성이 강조되고 있다.

2. 원헬스의 구성요소

원헬스의 구성요소는 인간과 동물의 건강, 그리고 환경의 3가지 요소로 이루어져 있다(그림 8-1-1).

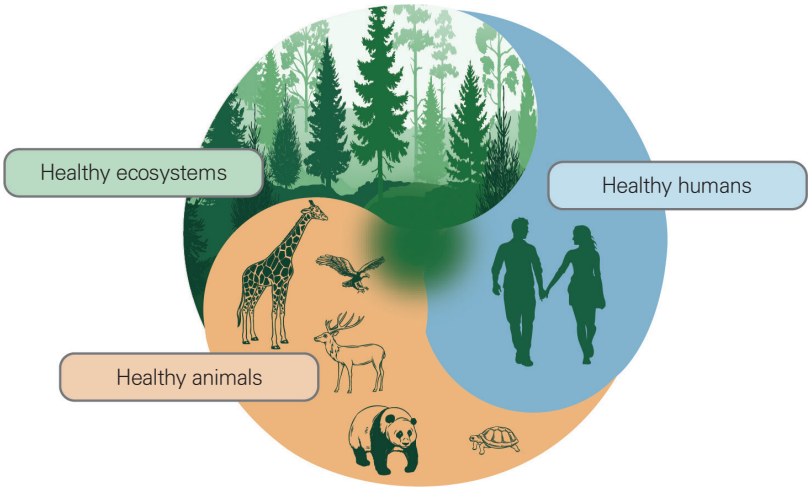


그림 8-1-1. 원헬스 구성요소

다음 사례를 들어 원헬스 구성요소에 대하여 설명하고자 한다.

1) 항생제내성

소, 돼지, 가금류, 염소를 포함한 국내 식량 생산 동물의 총 수는 약 2천만 마리이며, 2015년에만 95만 kg의 항생제가 축산동물에게 사용되었다. 축산 동물의 성장 촉진제로 사용되기도 하는 항생제는 노출 기간이 길어질수록 축산물 내 세균의 항생제내성을 증가시키며, 이는 이러한 육류를 섭취하는 사람에게 직접 또는 간접적으로 영향을 줄 수 있다. 따라서, 항생제내성 문제는 사람만의 문제가 아니며, 사람을 둘러싼, 동물(반려동물 포함)과 환경요소를 통합적으로 고려하는 원헬스 접근법이 무엇보다도 강조되고 있는 분야이다. 우리나라에서는 2016년 다부처가 참여하는 원헬스 항생제내성 관리정책을 내놓았고, 동물에서의 항생제 사용 규제 및 사용량 모니터링 체계를 구축하는 등 다양한 노력을 하고 있지만, 국내 인체 항생제내성 유병률은 여전히 높은 상황이며, 메티실린 내성 황색포도상구균 등과 같은 내성균의 경우 계속 증가세를 보이고 있다.

2) 조류인플루엔자

국내에서는 지속적으로 가금류, 야생조류에서의 조류인플루엔자 발생이 보고되고 있다. 최근 고양이 등의 포유류에서의 조류인플루엔자 감염이 보고되고 있어, 인간으로의 조류인플루엔자 확산이 우려되고 있다. 미국의 경우 농장종사자에서 조류인플루엔자 인체감염이 보고되는 등 기존과는 다른 발생 양상을 보이고 있어서 더욱 주의가 필요한 상황이다. 국내에서는 조류인플루엔자 인체감염을 예방하

기 위하여, 고병원성 조류인플루엔자가 발생한 농장대상으로 살처분 작업을 하고 있다. 또한 살처분 작업 중에 개인 보호 장비를 착용한 사람을 포함한 노출자에 대한 능동 감시와 유증상자 대상 항바이러스 예방적 투여가 이루어지고 있으며, 이는 조류인플루엔자의 인체감염을 막기 위한 원헬스적 관점에서의 관리 조치라고 볼 수 있다.

3) 레지오넬라

레지오넬라균은 토양에 상재하는 균으로써 인체감염의 경우 냉각탑 등 환경적인 요소를 통해 주로 감염되는 것으로 알려져 있다. 이전에 국내에서 보고된 레지오넬라증 사례의 노출 경로로는 공조 냉각탑, 월풀 스파, 수돗물 공급 시스템 등이 있다. 따라서, 레지오넬라 환례 발생 시 역학조사를 통하여, 가능한 노출경로(환경요인)에 대한 역학조사가 추가 환례 발생을 예방하는 것이 유행관리측면에서 매우 중요하다. 표 8-1-1은 환례 발생 지역 아파트 내 레지오넬라 균 배양을 통하여, 아파트 수계 시스템 내 감염원이 있음을 파악했던 자료이다. 레지오넬라 발생 관련 해당 아파트 주민들에 대한 레지오넬라 추가 발생 여부 감시와 아파트 수계 시설물에 대한 소독 및 고열수를 이용한 배관열처리 조치를 수행하였다.

표 8-1-1. 물검사 부위에 따른 온도, 잔류염소농도 및 레지오넬라 균증식 정도

물 검사 부위	온도	잔류염소농도(ppm)	레지오넬라 배양검사결과(CFU)
아파트 물탱크	20	0.43	0
아파트 수계 순환부	43	0.16	29,400
아파트 지하실 수도구	43	0.00	32,400
환례 거주지 내 수도구	43	0.04	72,000

4) 감염병 유행의 계절성

우리나라는 온대기후에 위치하고 있으며, 비교적 4계절이 구분되고 있는 상황으로 기후요인에 의해 감염병의 계절적 패턴을 보이는 경우가 많다(그림 8-1-2). 인플루엔자, 수족구병 등 감염병의 계절적 유행은 여러 병원체에서 확인되고 있으며 이는 장기적으로 기후변화에 따른 지역사회에서의 감염병 유행양상 또한 기존과는 다른 양상을 보일 우려가 있음을 보여준다.

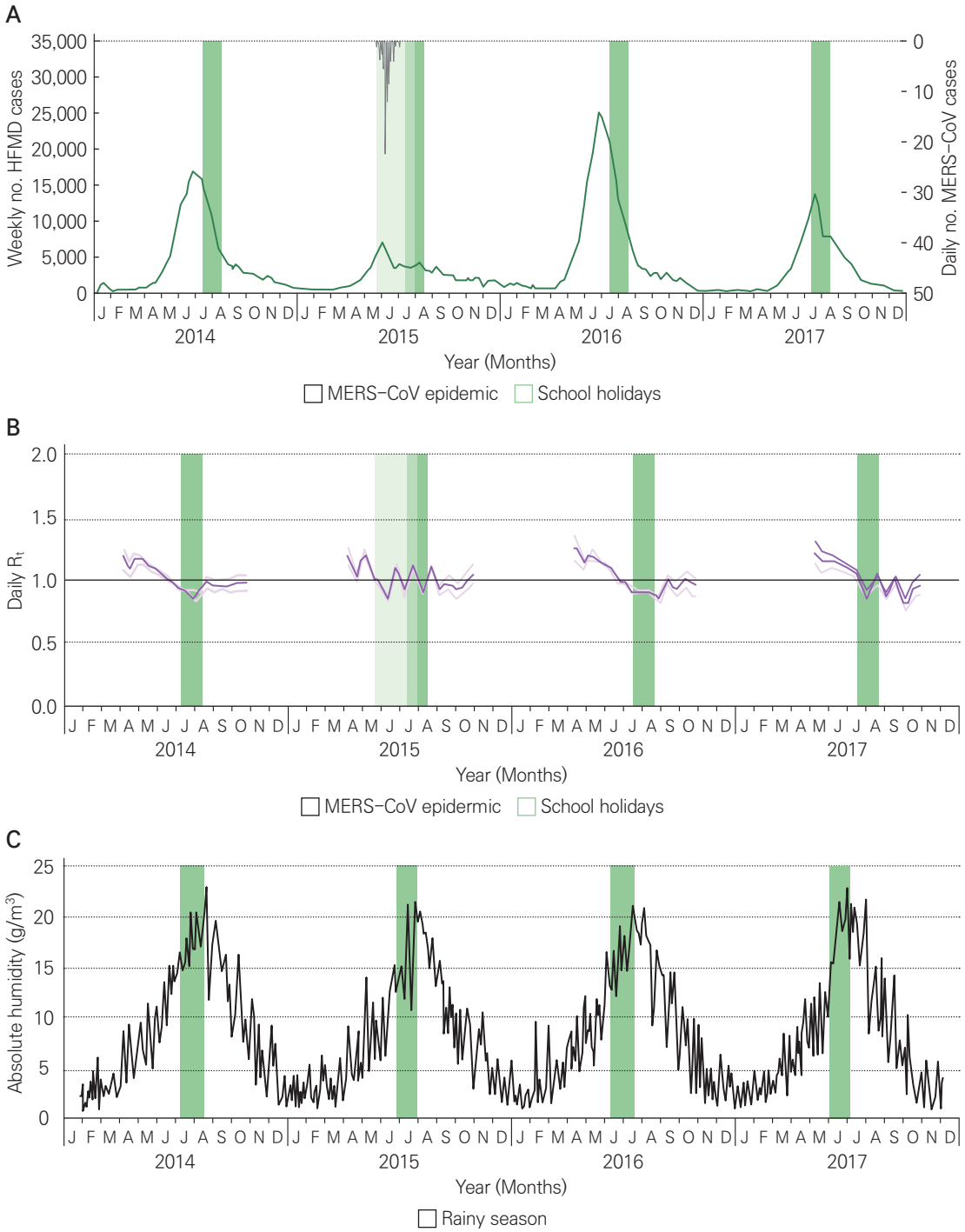


그림 8-1-2. 국내 2014년에서 2017년 사이 수족구병 환례 발생자 수 및 절대습도 변화 추이

Rt: 시간단위 t에 따른 환례1명당 다음 세대에 몇 명을 전파시키는지를 나타내는 유효재생산수

3. 원헬스의 중요성

우리가 전통적으로 알고 있는 인수공통감염병뿐만 아니라 현재 발생하고 있으며 앞으로 발생이 우려되는 대부분의 신종감염병이 동물로부터 유래하는 경우가 많고, 감염병의 특성을 고려할 때 원헬스적 접근이 요구되는 경우가 많다. 아직 원헬스적 접근이 분야간 정보공유 및 제한적인 연구차원에서만 머물고 있으나, 향후 정부와 학계, 국제사회와의 지속적인 협력을 통해서 이러한 접근을 더욱 확대해 나갈 필요가 있다.

참고 문헌

1. Ryu S, Han C, Ali ST, et al. Association of public health and social measures on the hand-foot-mouth epidemic in South Korea. J Infect Public Health. 2023;16(6):859-64.
2. Ryu S, Kim BI, Lim JS, et al. One Health Perspectives on Emerging Public Health Threats. J Prev Med Public Health. 2017;50(6):411-4.
3. Ryu S, Lim JS, Cowling BJ, et al. Low risk of avian influenza A (H5N6) transmission to depopulation workers in Korea. Influenza Other Respir Viruses. 2018;12(3):412-5.
4. Ryu S, Yang K, Chun BC. Community-acquired Legionnaires' Disease in a Newly Constructed Apartment Building. J Prev Med Public Health. 2017;50(4):274-7.
5. The Lancet. One Health: a call for ecological equity. Lancet. 2023;401(10372):169.
6. World Health Organization. Joint Tripartite (FAO, OIE, WHO) and UNEP Statement. 2024.

제2절 인수공통감염병의 원헬스적 접근

1 인수공통감염병 관리의 원헬스 적용과 활용

1. 인수공통감염병과 원헬스

원헬스(One Health)는 앞 장에서 본 것처럼 인간, 동물, 환경의 건강이 서로 밀접하게 연결되어 있다는 개념으로, 이 세 가지 영역의 건강을 통합적으로 고려해야 한다는 의미를 담고 있다. 이는 필수적으로 의학, 수의학, 생태학, 환경학 등 다학제적 접근을 요구한다.

인수공통감염병(zoonosis)은 하나의 생태계 내에서 동물과 인간 간 서로 전파되는 병원체에 의하여 발생하는 감염병으로 다른 어떤 분야보다 원헬스 접근법이 필요한 부분이다. 인수공통감염병의 예방과 관리에 원헬스 접근이 중요하다는 것은 비교적 오래전부터 논의가 이루어져 왔으며, 세계보건기구(WHO), 유엔식량기구(FAO), 세계동물보건기구(WOAH), 유엔환경계획(UNEP) 등 관련 국제기구에서도 원헬스는 인수공통감염병 대응을 위한 필수 전략으로 강조해 왔다. 특히 중증호흡기증후군(SARS, 2002~2003년), 신종플루(2009~2010년), 중동호흡기증후군(MERS, 2012년~현재), 코로나19(2019년~현재)와 같은 신종감염병의 대유행을 세계가 겪으면서 이들이 모두 인수공통감염병에서 출현했다는 점, 그리고 2024년 현재 유행하는 조류인플루엔자와 같이 또 다른 인수공통감염병이 사람 팬데믹의 시작이 될 것이라는 점에서 이제 원헬스 접근은 누구나 강조하는 중요한 개념이 되었다. 동물과 환경에서의 질병 발생을 조기에 감지하고, 선제적으로 대응하기 위해서도 원헬스 개념은 반드시 필요하다. 인수공통감염병에 대해서도 사람만을 대상으로 한 감시체계나 역학조사만으로는 이들을 예방하고 관리하는 것은 어렵다. 반면 동물과 환경에서의 유행의 관리는 사람에서 인수공통감염병의 예방에 아주 중요하다. 그러나 한편으로는 동물과 환경에서의 감염병 유행과 관리는 사람이 하는 작업이므로 이들을 같이 통합하여 관리하는 것이 필요하다.

2025년 3월 현재 감염병예방법상 정의된 인수공통감염병은 장출혈성대장균감염증, 일본뇌염, 브루셀라증, 탄저, 공수병, 동물인플루엔자인체감염증, 중증급성호흡기증후군(SARS), 변종크로이츠펠트-야

콥병(vCJD), 큐열, 결핵, 중증열성혈소판감소증후군(SFTS), 장관감염증(살모넬라균 감염증, 캄필로박터균 감염증) 등 모두 13종이다. 가축전염병예방법에는 이들 질병에 걸린 가축에 대한 신고를 받은 행정기관의 장은 즉시 질병관리청장에게 통보하도록 되어 있다.

그림 8-2-1은 1999년 1월에서 2015년 1월까지 우리나라 월별 렙토스피라증 발생률의 그래프이다. 여기서 렙토스피라증의 뚜렷한 계절성(주로 가을에 집중해서 발생)과 함께 점차 감소하는 추세를 볼 수 있다. 렙토스피라증은 위의 감염병예방법상 인수공통감염병에 포함되지는 않았지만, 쥐 등의 설치류가 자연숙주이고 쥐에서 배출된 균이 물이나 환경을 오염시키고 가축을 감염시키고 사람은 이들로부터 감염이 되는 인수공통감염병의 하나이다. 그림에서 보이는 붉은 선은 가축분뇨 관리에 관한 법률이 발효된 시점으로 이 이후에 유의하게 렙토스피라증 발생률이 감소하고 있는 것을 볼 수 있다. 즉, 가축 분뇨의 위생적 처리가 사람 렙토스피라증 감소로 이어진 것으로 이전에 사람을 대상으로 교육하고 홍보만 하던 것보다 동물에서의 감염원 관리가 효과적임을 보여준다.

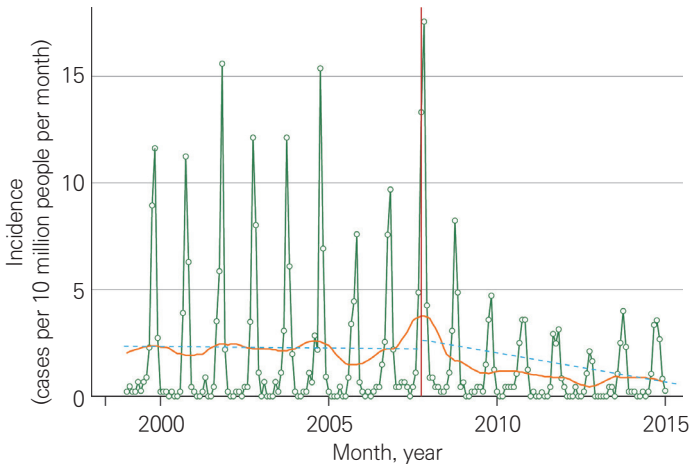


그림 8-2-1. 1999년에서 2015년까지 국내 렙토스피라증 월간 발생률
붉은 수직선은 가축분뇨관리에 관한 법률의 발효시점으로 이 시점 이후 국내 렙토스피라증이 유의하게 감소하였다.
[출처: Ryu, et al. 2017]

그림 8-2-2는 또 다른 인수공통감염병인 부르셀라증의 예이다. 2004년에서 2014년까지 소와 사람에서의 월별 부르셀라 발생률을 그린 것인데 소와 사람에서의 부르셀라증 발생률 증감이 동일한 양상을 보이고 있다 붉은 선은 소에서의 감시(수의분야에서는 예찰이라고 함)를 이전의 도축장에서만 수행하던 것을 농장에까지 확대한 시점이다. 이후 소에서의 부르셀라증이 감소하고 사람에서의 신고도 거의 소의 발생률과 함께 낮아지고 있음을 볼 수 있다. 즉, 소에서의 브루셀라 관리가 사람의 브루셀라증

예방과 관리에도 효과적으로 작용한다는 것이다.

이러한 사례들은 인수공통감염병의 관리에서 단순히 사람만 감시하고 진단, 치료하는 것이 매우 제한적인 접근법이며, 동물과 환경을 같이 관리하는 원헬스 접근의 중요성을 보여주고 있다.

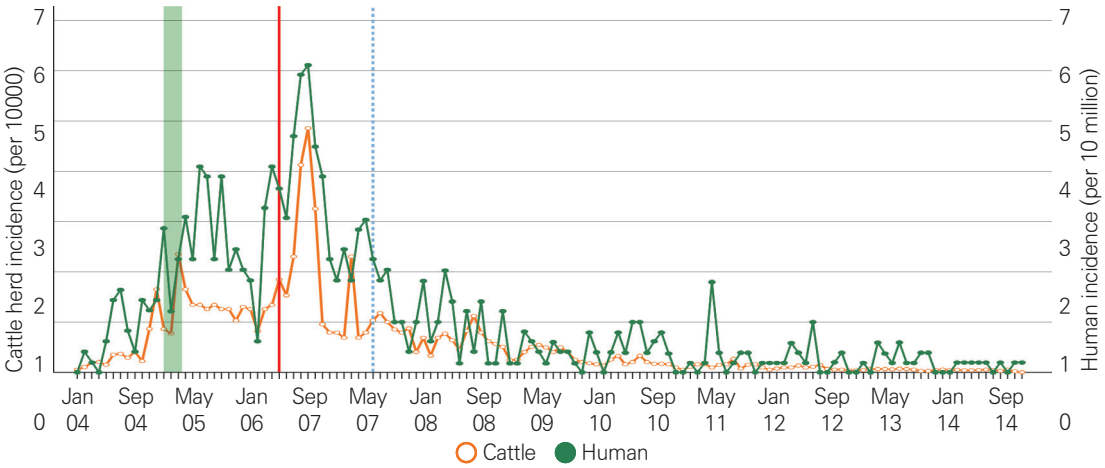


그림 8-2-2. 2004년에서 2014년까지 소(주황색 점)와 사람(초록색 점)의 월별 부르셀라증 발생률의 비교
붉은 수직선은 소에서의 부르셀라 감시를 도축장뿐 아니라 일반 농장에까지 확대한 시점이다.
[출처: Ryu S, et al. 2019]

또 다른 사례는 큐열이다. 2024년 질병관리청의 용역을 받은 고려의대 연구팀이 전남의 한 지역을 선정하여 소와 염소농장에서 근무하는 작업자 145명에 대해서 큐열항체를 조사하였다. 그 결과 12명(8.3%)이 급성큐열 확진, 15명(10.3%)이 급성큐열 추정, 14명(9.7%)이 만성큐열 추정으로 전체 양성률은 28.3%였다. 이때 개인 위험요인에 대한 다변량분석결과 축산업 종사 기간이 10~19년인 경우(10년 미만에 비해 OR=5.3, 95% CI: 1.4-21.5) 농장 내 소 또는 염소의 분만이 있었던 경우(OR=3.8, 95% CI: 1.2~13.9), 동물의 분변이 안구에 노출된 경우(OR=2.6, 95% CI: 1.0~6.7)가 유의한 위험요인이었고 이들은 큐열 예방에서 교육의 중요성을 보여준다고 할 수 있다. 하지만, 농장 위험요인에 대한 다변량분석에서 농장출입을 하는 차의 소독실시(OR=0.62, 95% CI: 0.08~0.97), 농장에 출입하는 사람을 소독하는 농장에서 사람의 큐열 양성률이 유의하게 낮았다(OR=0.21, 95% CI: 0.05~0.88). 농장을 출입하는 차량이나 사람의 소독은 가축전염병의 예방에서 중요하다고 알려진 것인데 이것이 사람에서의 큐열 예방에도 중요하다고 밝혀진 것이다. 그러나 통상 사람 환자에 대한 역학조사에서는 이렇게 동물의 위험요인을 조사하지 않기 때문에 이러한 점을 간과하게 된다.

2. 인수공통감염병에 대한 원헬스 감시체계

원헬스 감시체계(One Health Surveillance System)는 인간, 동물, 환경 분야의 질병 및 건강 자료를 공동으로 모니터링하고 분석하여 인수공통감염병 및 기타 건강 위협을 탐지, 예방 및 통제하는 통합적 접근 방식으로 정의할 수 있다. 원헬스 감시체계의 특징은 인간-동물-환경의 건강관련 자료를 결합하는 다부분 협업체계, 조기경보와 신속한 대응을 위한 실시간 질병모니터링, 병원-농장-검사실-연구기관을 연결하는 자료의 공유와 통합, 그리고 예방중심의 접근방식을 들 수 있다. 이러한 원헬스 감시체계의 개념적 정의는 아래의 그림 8-2-3처럼 표현할 수 있다. 그러나, 아직까지 이러한 원헬스 감시체계는 이상적이고 개념적인 측면이 강하다. 2018년에 수행된 한 체계적 고찰연구에서는 문헌상으로 보고된 전 세계 41개의 원헬스 감시체계에 대해서 분석을 했다. 이중에서 유럽, 캐나다, 미국에서 수행 중인 일부 모기매개감염병(웨스트나일열 등)에 대한 원헬스 감시체계(인간, 가축, 야생동물 및 매개모기 등에 대한 감시)가 통합적으로 잘 운영되고 있지만, 질병의 출현과 확산을 촉진하는 요인에 대한 정보는 포함되어 있지 않는 등 개념적 정의와는 여전히 차이가 있음을 기술하였다.

2019년 FAO, WHO, WOAH의 3개 기관이 작성한 3자간 인수공통감염병 가이드(Tripartite Zoonosis Guide, TZG)에는 국가가 원헬스 감시 및 정보 공유(Surveillance and Information Sharing, SIS)의 전략적 기술에 대한 역량을 개발하고 지원하는 것이 포함되어 있다. 이 3자 간 가이드(TZG)와 3개의 기관(FAO, WHO, WOAH)에서 개발한 감시 및 정보 공유 운영 도구(SIS OT)는 국가 당국이 인수공통감염병에 대한 다부문 감시 및 정보 공유를 수립하거나 강화할 수 있도록 지원하기 위해 개발된 것이다. 하지만, 이런 가이드와 도구의 제공에도 불구하고 개념적인 원헬스 감시체계를 현실화하는 것은 쉬운 일이 아니다. 원헬스 감시체계에 대한 진전을 보이고 있는 외국에서도 의학, 수의학, 환경분야의 각 부문 간 인식의 차이, 우선순위의 차이, 감시체계 구성요소에 대한 정확한 자료 습득의 어려움, 각 지표를 통합하는 분석 방법론의 복잡성, 실시간 자료공유가 안되는 문제 등등 다양한 문제를 경험하고 있다고 보고하고 있다.

2023년 원헬스 고위전문가 패널(One Health High-Level Expert Panel, OHHLEP)에서 원헬스 감시체계 개발에 대한 논의결과를 발표했는데, 그 결과는 원헬스 감시체계를 구축하는 6단계로 설명하고 있다. 이것은 감시체계에 참여자와 정책입안자가 원헬스 정의와 원헬스 감시범위에 동의하는 것부터 시작해서 단계별로 진행해야 한다는 것으로 이는 단시간 내에 쉽게 이루어지지 않는다는 것을 보여주고 있다. 첫 단계는 감시체계의 범위를 개발하는 것이고 두 번째 단계는 감시에 필요한 자료를 정의하는 것, 세 번째는 감시체계를 설계하고, 네 번째 단계는 감시체계의 거버넌스의 개발, 다섯 번째 단계는 통합된 운영 프로토콜의 개발, 여섯 번째 단계는 공동의 감시체계 로드맵을 개발하는 것이다.

현재 우리나라는 야생동물 질병 감시는 환경부 소관이고 가축 질병 감시는 농축산부 소관, 인간 감염병은 보건당국 소관이다. 인수공통감염병인 경우 농축산부에서 질병관리청으로 통보하고 있어서 자료의 공유가 안된다고 할 수는 없다. 또한 방역통합정보시스템을 활용하여 이러한 개념을 구체화할 수 있는 방안도 논의할 만하다. 우리나라도 인수공통감염병의 경우 이러한 통합적 원헬스 감시체계의 구축이 필요하기 때문에 현재 외국에서 진행되는 사례와 가이드들을 눈여겨 볼 필요가 있다.

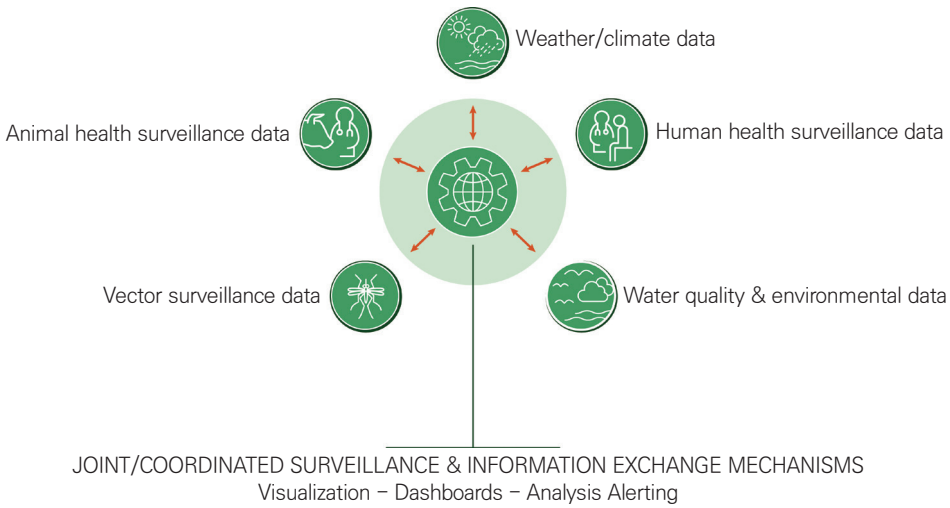


그림 8-2-3. 원헬스 감시체계에서 정보공유체계
[출처: FAO, WHO, OIE. Surveillance and Information Sharing Operational Tool. 2022]

3. 인수공통감염병에 대한 다자간 공동역학조사

우리나라의 경우 인수공통감염병의 유행 시 질병관리청과 농림축산검역본부에서 공동으로 역학조사를 실시할 수 있으며 이에 대한 매뉴얼이 공동으로 개발되어 있다(큐열, 부르셀라증, 결핵 등 주요 인수공통감염병).

큐열의 경우 사람 간 전파는 거의 없기 때문에 사람 큐열 감염자는 염소, 소, 양으로부터 감염되었다고 할 수 있다. 큐열은 가축뿐 아니라 야생동물과 진드기 등에서 발견되고 있기 때문에 전체 생태계에서 야생동물, 매개생물(진드기), 가축, 사람 등 원헬스의 각 부분을 고루 갖추고 있어서 원헬스 접근의 모형으로 활용하기 좋은 질환이기도 하다. 큐열의 경우 가축에서는 임신한 암컷에서 유산이나 사산이 발생하지만, 기타 동물에서는 거의 증상이 없어서 가축감염병 관리에서 우선순위가 낮고 발생신고 건수도 매우 적다. 그러나 사람에서는 급성과 만성큐열을 모두 일으킬 수 있고, 만성큐열의 경우는 발견

과 치료가 쉽지 않기 때문에 다른 인수공통감염병과 달리 사람에서 특히 중요한 인수공통감염병이라고 할 수 있다.

큐열 감염축이 발견되어 가축방역기관에 신고되면(KAHIS 보고체계), 이 정보는 질병관리청의 정보체계와 연동되어 공동으로 역학조사를 수행하게 되어 있다. 농장 근무자중에서 증상자에 대해서 병원 방문과 검사를 하도록 되어 있다. 만약 사람에서 먼저 신고가 되면 이 정보도 가축방역기관으로 통보가 되고 두 기관에서 공동으로 역학조사를 수행하는 체계이다. 하지만, 큐열의 경우 사람에서도 무증상감염이 많고, 가축에서 유사산 등으로 발견될 정도면 이미 농장에서는 상당수의 개체가 감염되어 있을 가능성이 높다. 따라서, 농장주나 농장 작업주의 경우는 이미 과거에 큐열에 감염되었거나 큐열인지 모르고 앓았을 가능성이 높다. 그러므로, 큐열 감염축이 신고되면 증상이 있는 사람에 대해서만 검사를 하는 것이 아니라 큐열 감염축이 나온 해당 농장이나 도축장 또는 육가공농장 종사자들에 대해서 모두 검사를 하고 역학조사를 하는 것이 좋다. 또한, 큐열의 경우 농장 주변의 야생동물이나 진드기에서도 발견되고 그 포자(spore)는 흙이나 먼지 등에서도 발견되어 호흡기로 들어오는 것이 주요 감염경로로 되어 있기 때문에 이러한 환경검체에 대한 검사도 필요하다.

공동역학조사가 실시되더라도 사람에 대한 조사와 농장 가축에 대한 조사 결과가 같이 통합되어 분석되어야 공동역학조사의 의미가 있을 것이다. 실제로 위에서 기술한 것처럼 가축의 위험요인이 사람에게도 위험요인으로 밝혀져 있다.

4. 인수공통감염병에 대한 원헬스 접근의 어려움

현실적으로 다부처, 다학문 영역이 하나의 지휘체계 아래 손발을 맞추어 사업을 하는 것은 쉽지 않다. 원헬스 접근이 원활하게 이루어지기 위해서는 아래와 같은 어려움이 해결되어야 한다.

1) 원헬스 거버넌스 체계의 미흡 및 부처 간 협력 강화의 어려움

- 보건복지부, 환경부, 농림축산식품부와 이들 각 부 산하의 여러 기관이 관여하지만, 부처 간 공동사업을 위한 협력은 아직 원활하지 않은 부분이 있다.
- 원헬스를 추진할 수 있는 명확한 거버넌스 체계가 부족하고 각 기관의 법적·행정적 권한이 분리되어 있어 공동 대응이 어렵다.

2) 통합 감시체계 구축의 어려움

인간·동물·환경의 질병 감시 자료가 통합적으로 관리되지 않고, 기관별로 수집된 데이터가 다른 기관

과 효율적으로 공유되고 분석되는 체계가 미흡하다. 원헬스 관점에서 야생동물, 가축, 사람에서의 감염병 감시정보가 같이 공유되고 분석되는 감시체계가 바람직하다.

3) 산업적 이해관계 충돌

각종 개발사업, 축산업, 농업, 환경 보호 정책 간의 이해관계가 대립되는 부분이 있다. 예를 들면 집약적 축산업(공장식 축산)은 가축 감염병 확산 위험을 높이지만, 경제적 이익 때문에 규제 강화가 어려운 부분이 있고, 각종 개발 사업(삼림 벌채, 도시 확장 등)과 자연보호 간의 충돌이 있을 수 있다.

4) 학문 간 협력 부족 및 인식 부족

의학 및 공중보건학, 수의학, 생태학, 환경과학 간의 다학제적 협력이 충분하지 않다. 많은 보건학 및 수의학 전문가들이 원헬스의 중요성은 인식하고 있으나 문제해결에 있어서는 자신의 분야 중심으로 문제를 바라보는 경향이 있으며, 의과대학, 수의과대학 및 공중보건학 교육 과정에서 원헬스 개념이 충분히 다루어지고 있지 않다. 다학제적 연구 및 감시 활동을 수행할 현장 전문가(공중보건학자, 수의학자, 환경학자)를 충분히 배출하지 못하고 있다.

5) 원헬스 연구 및 정책의 단기적 특성

인수공통감염병에 대한 원헬스 접근은 장기적 관점에서 지속적인 감시와 연구가 필요하지만, 다른 분야와 유사하게 단기적인 성과 중심의 연구 및 정책이 많아 원헬스에서 요구되는 장기적 접근을 어렵게 한다.

이 외에도 원헬스 수행의 어려움으로 국제적 협력의 어려움을 들 수 있는데 WHO, FAO, OIE 등 국제기구에서는 원헬스를 강조하지만, 각국의 대응 속도가 다르며, 선진국과 개발도상국 간 감염병 대응 역량의 격차가 큰 것들이 걸림돌이 되고 있다.

원헬스 접근과 시행에서 이러한 어려움은 비단 우리나라만 있는 것은 아니고 각국에서 유사한 경험을 하고 있다. 우리나라는 관련 전문가들로 위원회를 만들어 운영하고 정기적인 포럼을 개최하면서 다학제 간 상호 이해를 증진시키고 있다. 주요 인수공통감염병의 경우 사람과 동물의 공동역학조사 지침을 만들어 공동대응할 수 있는 방안을 제시하고 있고 특히 원헬스 전략을 제3차 감염병기본계획(2023~2027년)의 주요한 접근방법으로 포함하여 꾸준한 원헬스 접근이 강화된 관리 정책을 펼치고 있다.

5. 맺음말

원헬스 활성화는 다부처 간 협력, 재정 및 인력 지원, 데이터 공유, 다학제적 접근이 필수적이다. 하지만, 행정적, 경제적, 학문적 한계로 인해 더디게 활성화가 진행되는 상황이다. 따라서, 국가 및 국제적 협력을 바탕으로 장기적인 원헬스 감시 시스템 구축, 다학제적 연구 지원, 원헬스 역학조사의 활성화와 같은 정책적 개선이 지속되어야 한다.

2 인수공통감염병 유행시 동물 및 환경보건

1. 국내 주요 인수공통감염병 현황 및 공동역학조사 관련 법령

1) 국내 감염병예방법과 가축전염병예방법상 공통 관리대상 인수공통감염병

① 감염병의 예방 및 관리에 관한 법률에서 인수공통감염병으로 관리하고 있는 질병 중에서 가축전염병예방법상에서 법정 전염병으로 지정된 감염병(전염병)은 8개로 탄저, 동물인플루엔자, 우결핵, 일본뇌염, 브루셀라증, 공수병, 변종 크로이츠벨트-야콥병, 큐열이 있다(표 8-2-1).

- 8개 감염병 중, 변종 크로이츠벨트-야콥병을 제외하고, 모두 국내 가축에서 발생사례가 존재하며, 특히 감염병예방법상 1급에 해당하는 탄저의 경우 가축에서 1978년 제주도에서 최초 보고 된 이후, 최근 2008년 영천에서 1건이 발생한 바 있다.

표 8-2-1. 국내 인수공통감염병의 법령 지정 현황

감염병	등급		보고대상 여부 (인체/가축)
	감염병예방법	가축전염병예방법	
탄저(Anthrax)	1급	2종	인체: 보고대상 가축: 보고대상
중증급성호흡기증후군(SARs)	1급	미지정	인체: 보고대상 가축: 보고대상 아님
동물인플루엔자 인체 감염증 [Animal influenza (Avian, Swine)]	1급	1종(고병원성조류 인플루엔자), 2종(돼지인플루엔자), 3종(저병원성조류인플루엔자)	인체: 보고대상 가축: 보고대상

감염병	등급		보고대상 여부 (인체/가축)
	감염병예방법	가축전염병예방법	
우결핵(M. Bovis)	2급	2종	인체: 보고대상
			가축: 보고대상
장 출혈성대장균 감염증 (Enterohemorrhagic E coli)	2급	미지정	인체: 보고대상
			가축: 보고대상 아님
일본뇌염(Japanese encephalitis)	3급	2종	인체: 보고대상
			가축: 보고대상
브루셀라증(Brucellosis)	3급	2종	인체: 보고대상
			가축: 보고대상
공수병(Rabies)	3급	2종	인체: 보고대상
			가축: 보고대상
변종 크로이츠펠트-야콥병 (vCJD)	3급	2종[소해면상뇌증(BSE)]	인체: 보고대상
			가축: 보고대상
큐열(Q-fever)	3급	2종	인체: 보고대상
			가축: 보고대상
중증열성혈소판감소증후군(SFTS)	3급	미지정	인체: 보고대상
			가축: 보고대상 아님

- 8개의 질병 중, 가축에서 능동예찰(Active Surveillance, 능동감시) 체계가 구축되어 있는 질병은 동물인플루엔자인체감염증에 해당하는 고병원성조류인플루엔자, 저병원성조류인플루엔자(혈청예찰 실시), 우결핵, 일본뇌염, 브루셀라증이 있으며, 큐열의 경우 2025년부터 혈청예찰(항체 검사)을 실시할 예정에 있다.
  * 야생생물보호 및 관리에 관한 법률(야생생물법)에 따라, 조류인플루엔자, 야생포유류 결핵, 야생포유류 브루셀라병, SFTS, 야생동물 광견병에 대해서, 질병 발생 현황을 공개하고 있음

② 가축 역학조사대상 질병

- 위의 8개 인수공통감염병 중, 가축전염병 예방법 시행규칙 제15조(역학조사의 대상 등)에 따라 농식품부 및 지자체에서 역학조사를 실시해야 하는 질병은 고병원성조류인플루엔자이며, 농림부 위생검역과-51580-165(2002.2.5.)에 따라서 탄저, 브루셀라병, 결핵병, 광견병에 대해서 시도 가축방역기관에서 역학조사를 실시하여 한다.
- 따라서, 8개 중 위의 5개 질병(고병원성조류인플루엔자, 탄저, 브루셀라병, 결핵병, 광견병)에 대해서는 역학조사를 실시하고 있으며, 이는 사람과 동물의 질병의 추가확산 및 관리를 위해

서 원헬스적 목적에 따라, 공동역학조사가 가능하다고 볼 수 있다.

③ 공동역학조사 관련 법령 및 제도

- 위에서 언급한 원헬스를 위한 인수공통감염병에 대한 역학조사는 각 부처별로 감염병예방법(질병관리청), 가축전염병 예방법(농식품부), 야생생물법(환경부) 같은 법적 근거에 따라 이루어져야 한다.
- (사람) 감염병예방법 제18조(역학조사) 1항, 제18조의2(역학조사의 요청)

감염병예방법 제18조(역학조사) 1항

감염병예방법 제18조(역학조사) 1항

질병관리청장, 시·도지사 또는 시장·군수·구청장은 감염병이 발생하여 유행할 우려가 있거나, 감염병 여부가 불분명하나 발병 원인을 조사할 필요가 있다고 인정하면 지체없이 역학조사를 하여야 하고, 그 결과에 관한 정보를 필요한 범위에서 해당의료기관에 제공하여야 한다. 다만, 지역확산 방지 등을 위하여 필요한 경우 다른 의료기관에 제공하여야 한다.

제18조의2(역학조사의 요청)

감염병예방법 제18조의2(역학조사의 요청) 1항

「의료법」에 따른 의료인 또는 의료기관의 장은 감염병 또는 알수 없는 원인으로 인한 질병이 발생하였거나 발생할 것이 우려되는 경우 질병관리청장 또는 시·도지사에게 제18조에 따른 역학조사를 실시할 것을 요청할 수 있다.

- (가축) 가축전염병 예방법 제13조(역학조사) 시행규칙 제15조(역학조사 대상 등)

가축전염병 예방법 제13조(역학조사)

가축전염병 예방법 제13조(역학조사) 1항

국립가축방역기관장, 시·도지사 및 시·도 가축방역기관장은 농림축산식품부령으로정하는 가축전염병이 발생하였거나 발생할 우려가 있다고 인정할 때에는 지체 없이 역학조사를 하여야 한다.

시행규칙 제15조(역학조사 대상 등)

가축전염병 예방법 시행규칙 제15조(역학조사 대상 등) 1~4항

그 밖의 가축전염병 중 농림축산식품부장관 또는 검역본부장이 역학조사가 필요하다고 인정하는 가축전염병.

검역본부장, 시·도지사 및 시·도 가축방역기관장은 제1항 각 호에 따른 가축전염병이 발생하였거나 발생할 우려가 있다고 인정하는 경우에는 지체없이 다음 각 호의 구분에 따라 역학조사를 실시하여야 한다. 이 경우 가축전염병의 방역을 위하여 긴급한경우에는 검역본부장, 시·도지사 및 시·도 가축방역기관장이 공동으로 실시하여야 한다.

시·도지사 및 시·도 가축방역기관장이 제2항제2호에 따라 역학조사를 하는 때에는 그 결과를 검역본부장에게 제출하여야 하고, 검역본부장은 역학조사를 추가로 실시하여야 할 필요가 있다고 인정되는 경우에는 시·도지사 또는 시·도 가축방역기관장에게 추가로 역학조사를 하게 할 수 있다.

- (야생동물) 야생생물법 제34조의9(역학조사)

야생생물법 34조의9(역학조사)

야생생물법 34조의9(역학조사)

국립야생 동물질병관리기관장과 시·도지사는 다음 각 호의 어느 하나에 해당하는 경우 원인규명 등을 위한 역학조사(疫學調査)를 할 수 있다.

1. 야생동물 질병이 발생하였거나 발생할 우려가 있다고 인정한 경우
2. 야생동물에 질병 예방 접종을 한 후 이상반응 사례가 발생한 경우
3. 시·도지사(국립야생 동물질병관리기관장에게 요청하는 경우에 한정한다) 또는 관계 중앙행정기관의 장이 요청하는 경우

누구든지 국립야생 동물질병관리기관장 또는 시·도지사가 제1항에 따른 역학조사를 하는 경우 정당한 사유 없이 이를 거부 또는 방해하거나 회피해서는 아니 된다. 〈개정 2019. 11. 26.〉

제1항에 따른 역학조사의 시기 및 방법 등에 관하여 필요한 사항은 환경부령으로 정한다.

2. 동물단계 인수공통감염병 역학조사 방법 소개

1) 역학조사 방법

(1) 역학조사 절차

① 가축(산업동물)에서 역학조사는 병원체에 특성에 따라 조사 대상 및 조사 기간 등[예: 조류인플루엔자 경우 부화장, 구제역 가축시장(우시장) 등]이 일부 다를 수 있지만, 역학조사 실무 매뉴얼에

따라, 6단계를 거쳐 실시한다(그림 8-2-4, 5).

- 1단계: 기초정보 수집 및 분석
- 2단계: 신고(발생) 농장 현장 역학조사 및 추적조사
- 3단계: 유입원인 및 전파에 가설 설정
- 4단계: 유입원인 및 전파 가설에 대한 검증
- 5단계: 역학조사 결과에 대한 검증 및 자문
- 6단계: 역학조사보고서 작성

② 가축에서 역학조사는 사람의 역학조사와 달리, 직접 환축과 의사소통을 할 수 없기 때문에, 가축 소유주 및 농장 종사자 등과 인터뷰를 통해서 주로 실시하고 있으며, 다양한 정보와 자료를 수집하여 실시한다. 다음은 주로 발생 농장에 도착하여 현장조사 시 수집 사용하는 자료 및 정보 출처이다.

- 축산업등록증 및 허가증(농장 면적 및 구조)
- CCTV(유입경로 파악, 방역 위반 사항 점검 등)
- 사양일지(폐사두수, 폐사일지, 약품 사용 내역 등)
- 소독실시기록부(일자별 소독 실시 내역 등)
- 출입기록부(일자별 농장 출입자 내역 등)
- 거점소독필증(농장 출입 시, 출입차량의 소독여부 유무 파악)

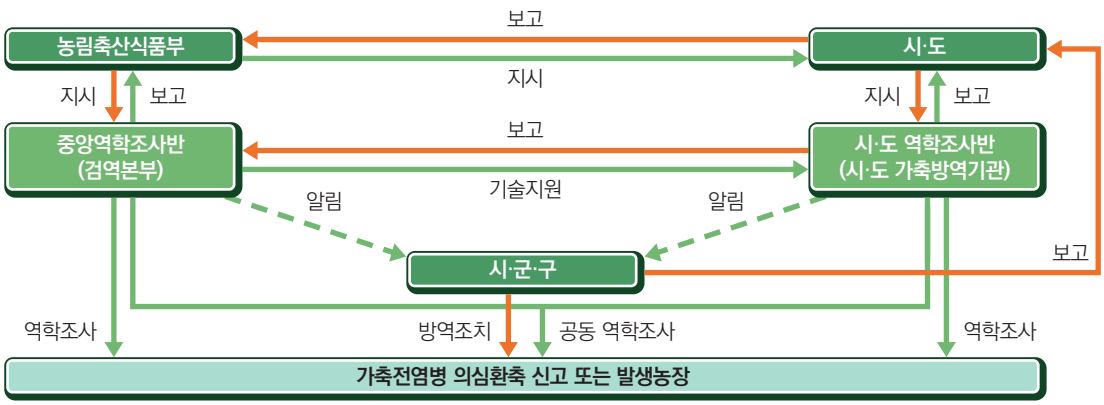


그림 8-2-4. 기관별 역학조사 수행 흐름도
[출처: 고병원성 조류 인플루엔자 실무메뉴얼, 농림축산검역본부, 2015.]

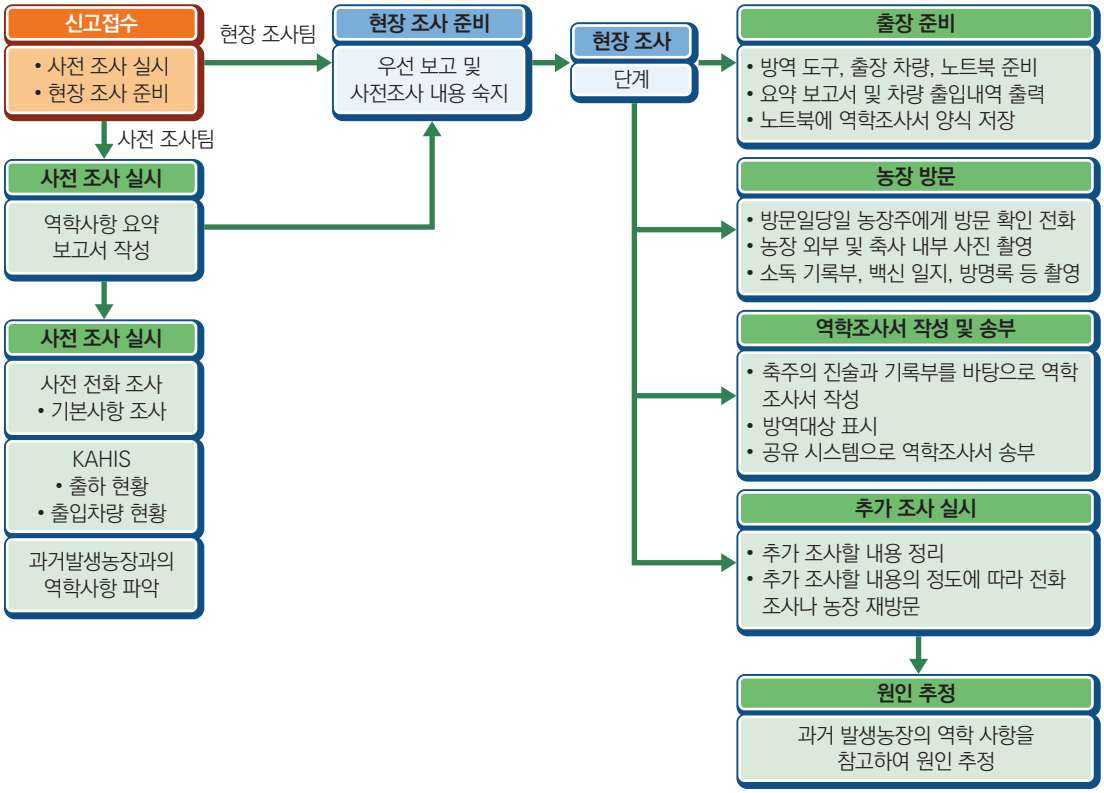


그림 8-2-5. 가축에서 역학조사 수행 및 과정 흐름도

(2) 역학조사 시 주요 사용 데이터

① 가축(산업동물)에서 역학조사 시, 농장 내 자료 이외에 사전에 다양한 정보 출처를 통해서 아래와 같은 데이터를 수집하여 역학조사(추적조사 및 유입경로 파악 등에)에 활용하고 있다(표 8-2-2).

표 8-2-2. 가축 역학조사 시 주요 사용 데이터 종류

항목	설명	특징	출처
축산차량 출입 데이터	• 가축전염병예방법 제17조의3에 따라 축산차량 등록 및 GPS 단말기 장착한 차량(시설출입차량)이 축산관계시설에 방문 정보	• 차량종류별, 축산관계시설별로, 방문시간 파악 가능 • 차량 이동동선을 파악 가능	농림축산검역본부 국가가축방역 통합시스템

항목	설명	특징	출처
축산차량 출입 데이터	• 축산관계시설에 출입하는 차량(소유자 및 운전자) – 가축·원유·알·동물약품·사료·조사료·가축분뇨·퇴비·왕겨·쌀겨·톱밥·깔짚·난좌·가금부산물 운반, 진료·예방접종·인공수정·컨설팅·시료채취·방역·기계수리– 가금 출하·상하차 등을 위한 인력운송 – 가축사육시설의 운영·관리(가축전염병예방법 제17조제1항제1호에 해당하는 자가 소유하는 차량의 경우에 한정한다) • 축산관계시설 가축사육시설, 도축장, 집유장, 식용란 수집판매업소, 사료공장, 가축시장, 가축검정기관, 종축장, 부화장, 비료공장, 가축분뇨 처리업체		농림축산검역본부 국가가축방역 통합시스템
축산물이력 데이터	• 소이력번호 부여됨에 따라, 개체별 농장간 이동 및 출하 정보	• 개체별 출생 및 양수, 양도, 도축 일자 파악 가능 • 구제역 백신접종일장, 브루셀라 검사 일자 파악 가능	축산물품질관리원 이력관리시스템
도축 출하 데이터	• 농장의 가축의 도축장 출하 정보	• 농장별 출하일자 및 출하두수 파악 가능	농림축산검역본부 국가가축방역 통합시스템
예찰 데이터	• 농장별 주요 가축전염병에 대한 예찰 정보	• 과거 가축전염병 검사 일자 및 결과 파악 가능	농림축산검역본부 국가가축방역 통합시스템
방역점검 데이터	• 농장별 주요 소독 및 차단 방역점검정보	• 농장별 소독 및 차단방역 위반사항 파악 가능	농림축산검역본부
국경검역 데이터	• 농장종사자 해외 출입국 정보	• 주요 가축전염병 발생국 방문 정보 파악 가능	농림축산검역본부

2) 역학조사 세부 주요 특징 및 유의사항

가축전염병에 대한 역학조사는 가축전염병 등급에 따라, 그 강도와 조사 내용이 다르다. 예를 들어, 1종 가축전염병인 고병원성 조류인플루엔자는 발생일자(신고일자) 기준 과거 21일간 조사를 실시하고, 이러한 조사 결과를 바탕으로 역학관련 농장 및 출입차량에 대한 이동제한 조치를 취하고 있다. 반

면에 2종 가축전염병 소브루셀라, 소결핵 등의 경우에는 양도(판매한 농장)에 대해서 조사를 실시하여 역학관련 농장이라고 보고, 해당 농장에 검사를 실시하나, 역학관련 농장에 대해서 별도의 이동제한 조치를 취하고 있지 않다. 자세한 사항은 조류인플루엔자 긴급행동지침(농식품부, 2022.3.25.), 결핵병 및 브루셀라병 방역실시요령(농식품부 고시, 2024.2.1.)에 나와 있다.

3. 인수공통감염병 동물 및 환경분야 협력 방법

1) 사람-동물의 원헬스적 협력 방법

각 부처의 해당 법령의 설립 취지와 이에 따른 역학조사 목적이 상이할 수 있기 때문에 상호 간의 차이를 이해하고, 소통하는 것이 공동역학조사를 성공적으로 수행하는 데 가장 중요한 요소라고 할 수 있다.

아래에서 보는 것과 같이, 감염병예방법의 목적은 국민 건강의 증진 및 유지인데 반해, 가축전염병 예방법의 경우에는 축산업 발전이 주 목적이고, 야생생물법의 경우에는 야생생물의 멸종을 예방, 생물다양성 증진, 건전한 자연환경을 확보가 목표이기 때문에, 이 목적을 이룰 수 있는 감염병(전염병)에 대해서 우선순위를 두고, 예방 및 관리, 이를 위한 역학조사를 실시하고 있다. 따라서, 각 부처마다의 우선순위 대상 질병이 다를 수 있고, 이로 인해서 부처마다 질병에 대한 감시체계 구성, 관리 강도가 다를 수 있다. 따라서 각 부처의 공통의 아젠다를 모색하는 것이 필요하다.

- 감염병예방법 목적: 국민 건강에 위해(危害)가 되는 감염병의 발생과 유행을 방지하고, 그 예방 및 관리를 위하여 필요한 사항을 규정함으로써 국민 건강의 증진 및 유지에 이바지
- 가축전염병 예방법 목적: 가축의 전염성 질병이 발생하거나 퍼지는 것을 막음으로써 축산업의 발전, 가축의 건강 유지 및 공중위생의 향상에 이바지
- 야생생물법 목적: 야생생물과 그 서식환경을 체계적으로 보호·관리함으로써 야생생물의 멸종을 예방하고, 생물의 다양성을 증진시켜 생태계의 균형을 유지함과 아울러 사람과 야생생물이 공존하는 건전한 자연환경을 확보

2) 사람-동물의 역학조사 분야에서 협력 방법

(1) 공동역학조사 매뉴얼 개발

현재 사람-동물에 대해서 공동역학조사 매뉴얼이 아래와 같이 개발되어 있다. 이러한 역학조사 매뉴얼을 기반으로 조사 대상 방법을 표준화하고, 향후 수집된 데이터를 바탕으로 정책 수립을 위한 과학적 근거를 제시할 수 있다.



큐열 공동역학조사 매뉴얼



브루셀라병 공동역학조사 매뉴얼

그림 8-2-6. 큐열 및 브루셀라증(병) 공동역학조사 매뉴얼

(2) 통합 데이터 분석 방법 개발

- 수집된 데이터를 바탕으로 다음과 같은 방식으로 원헬스적 접근이 가능함. 예를 들어, 발생농장과 비발생농장의 주요 요인에 대해 조사를 실시하고, 이를 통해서 농장의 유입경로 및 위험요인이 파악이 가능하고, 발생농장 중에서 감염된 사람과 비감염된 사람의 방역수칙 준수를 비교를 통해서, 위험한 농장에서 사람으로의 감염경로 및 위험요인을 파악함으로써, 사람에게 미치는 영향 및 위험을 파악할 수 있다.
- 추가적으로 발생농장의 출입하여, 감염개체와 접촉한 사람들[수의사, 가축 운반자(상하차반), 인공수정사, 가축분뇨운반자 등]에 대한 조사를 통해서 이러한 예방 방안 마련에 기여할 것으로 생각된다.

표 8-2-3. 농장 및 사람의 역학 분석 예시

위험요인	큐열 발생(양성) 가축농장(30호)		큐열 비발생(음성) 가출 농장(60호)	
	큐열 감염 농장 근로자	큐열 비감염 농장 근로자	큐열 감염 농장 근로자	큐열 비감염 농장 근로자
농장 근로자 마스크 미착용	50명	10명	10명	110명

(3) 상호 연결성 및 통합 위험도 분석

사람과 동물의 원헬스적 예방 관리를 위해서는 우선적으로 동물 단계에서 위험도를 분석하는 것이 선행되어야 한다. 다만 이와 더불어 매우 중요한 부분이 사람과 동물 간의 상호 연결성(interconnectivity)을 파악하고 이해하는 것이 매우 중요하다. 즉 사람-가축, 사람-야생동물, 사람-반려동물 간의 접촉 방식 및 관계를 파악하고 이러한 관계가 외부요인에 따라 어떻게 달라지는지를 파악함으로써 사람에게 미치는 영향 및 위험을 파악할 수 있다.

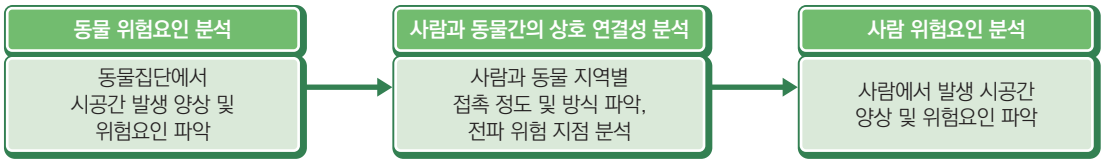


그림 8-2-7. 상호 연결성 및 통합 위험도 분석

참고 문헌

1. 박성대, 유대성. 위험 매트릭스(Risk Matrix)를 활용한 조류인플루엔자 인체감염증위험지역 평가. 한국가금학회지. 2023;50(1):41-50.
2. 질병관리청, 농림축산검역본부. 인수공통감염병 공동 역학조사 매뉴얼_브루셀라증(병). 2023.
3. 질병관리청, 농림축산검역본부. 인수공통감염병 공동 역학조사 매뉴얼_큐열. 2023.
4. 농림축산검역본부, 유대성. 고병원성조류인플루엔자 실무매뉴얼. 2015. 16-9쪽.

제3절 항생제내성 관리와 원헬스

1 항생제

감염과 관련된 역사에서 세균설(germ theory) 못지않게 중요한 전환점은 항생제의 도입이다. 19세기 말 경부터 다시금 재평가받으며 본격적으로 다뤄지기 시작한 세균설은 항생제의 역사와 그 궤적을 같이 한다. 당시 Louis Pasteur는 세균들끼리 주고 받는 길항작용의 본질을 알아낼 수 있다면 세균 질환 치료의 해결책을 마련할 수도 있다는 의견을 제시하였으며, 이는 항생제가 본격적으로 발견되기 전부터 이미 항생(antibiosis) 개념을 도입하고 응용할 수 있다는 혜안을 보여준 것이었다. 실제로 1877년 Pasteur와 Robert Koch는 어떤 공기 매개 세균이 탄저균을 억제하는 현상을 보고하기도 하였다. 이렇게 미생물이 내는 물질로 다른 미생물을 억제한다는 개념의 연구가 시작되기도 했지만, 다른 한편으로 화학물질을 합성해서 항생제를 개발하는 연구를 이끄는 사람들도 있었다. 대표적으로 독일의 Paul Ehrlich는 19세기 말부터 일부 아닐린 염료 물질이 생물 종에 따라 다르게 착색하는 결합 선택성이 있다는 사실을 바탕으로 세균은 착색하지만 사람의 세포에는 결합하지 않는 화학 물질을 개발하면 항생제로 쓸 수 있겠다는 가설을 세우고 개발을 시작하였다. 이와 같은 다양한 연구들의 성과물들이 축적되고 이를 상품화하려는 노력이 결합되면서 20세기 초반 드디어 항생제의 시대가 시작되었다.

항생제(antibiotics)는 antibiosis에서 유래된 용어이며, 흙 속에 사는 세균을 연구하다가 streptomycin을 발견하여 결핵의 새로운 치료지평을 열고 노벨상을 수상했던 Selman A. Waksman에 의해 1942년 처음 사용되었다. 이는 미생물이 내는 물질로, 다른 미생물의 성장을 억제하거나 죽이는 기능이 있는 모든 것을 통칭하였다. 그러나, 넓은 의미에서는 미생물이 만들어내는 물질 뿐만 아니라 화학적으로 합성된 약물, 혹은 기존 항생제 구조를 일부 변형하여 만든 반합성 약물 등도 모두 항생제에 포함시킨다. 우리는 흔히 항생제를 세균을 겨냥한 물질로 사용하고 있지만, 항바이러스제, 항진균제, 항기생충제 등과 구별하기 위해서는 항균제라는 용어가 더 정확하다고 할 수 있다. 따라서, 항균제, 항바이러스제, 항진균제, 항기생충제 등을 모두 포괄하여 항미생물 제제(antimicrobial agents)라고도 한다.

2 항생제내성

1. 항생제 이전 시대

항생제 이전 시대(pre-antibiotic era)는 '인류가 항생제를 보유하게 되기 이전까지의 시대'라고 할 수 있다. 항생제를 '발명'한 것이 아닌 '보유'하게 된 것으로 표현한 이유는 항생제 또는 항생 물질은 만들어진 것이 아니고 '찾아낸' 것이기 때문이다. 앞서 설명한 것처럼 항생물질은 미생물이 지구상에 나타난 시점부터 이미 존재하고 있었으며, 수많은 미생물들이 각자의 세력을 구축하고 생존 경쟁을 하면서 서로간에 다양한 견제 수단을 활용했는데, 항생 물질은 이런 수단 중 하나였다. 또한, 수없이 많은 항생물질에 대항하는 방어기전들도 생겨나게 되는데, 여기에 해당하는 것이 바로 항생제내성이다.

2. 항생제의 시대

인류는 항생제를 본격적으로 보유하기 위해 20세기 초까지 기다려야 했다. 1928년 플레밍이 페니실린을 분비하는 곰팡이를 발견하면서 시작된 항생제 시대의 개막은 현대의학에 여러 가지 변화를 불러왔다. 이전까지 인류의 생명에 최대의 위협이 되었던 감염증을 치료할 수 있게 되었을 뿐만 아니라 외과의 큰 수술, 항암요법, 나아가 각종 장기이식에 이르는 여러 가지 새로운 치료법이 가능하게 되었다. 그러나, 이러한 "기적의 약(항생제)"도 지난 세기 동안 오·남용되며 중대한 고비를 맞고 있다. 즉, 항생제내성의 출현과 급속한 확산, 새로운 항생제 개발의 지연 등으로 항생제 이전의 시대와 같은 위험한 공백상태("항생제 시대의 종말")가 초래되고 있는 것이다. 항생제내성의 발현과 전파는 나라에 따라 정도의 차이가 있을 뿐 전세계적으로 커다란 문제가 되고 있다. 일반적으로 항생제내성은 모든 항미생물제에 대해 생길 수 있는데, 여기에서는 주로 항균제내성을 중심으로 설명한다.

3. 항생제내성

1) 정의

항생제내성이라는 용어는 널리 사용되고 있지만 명확한 정의가 내려져 있지 않고 국가 간의 기준도 차이가 있을 수 있다. 미국의 CLSI (Clinical Laboratory Standards Institute)는 치료 용량의 항균제를 투여하였을 때 세균이 일으킨 감염병을 적절하게 치료할 수 있는 경우를 감수성(susceptible)으

로, 치료 용량의 항균제 투여 후 얻을 수 있는 항균제 농도로 세균을 억제할 수 없는 경우를 내성(resistant)이라고 정의한다. 검사실에서는 각 균주별로 항균제의 종류에 따라 세균의 성장을 억제하는 최소 억제농도(minimum inhibitory concentration, MIC)가 정해져 있기 때문에 이를 기준으로 감수성과 내성을 판단한다.

2) 자연내성 혹은 획득내성

항균제내성은 유전적 특성에 의해 처음부터 듣지 않는 자연내성(intrinsic resistance)과 원래는 감수성이었던 세균이 돌연변이(mutational resistance) 혹은 다른 세균에서 유전자를 획득(transferrable resistance)하여 내성을 나타내는 획득내성(acquired resistance)으로 나눌 수 있다. 전자의 예로는 vancomycin의 경우 거의 대부분의 그람양성균에는 항균력이 있으나 분자량이 커서 그람음성균의 외막을 통과하지 못하기 때문에 그람음성균에는 듣지 않는다. 반면에 penicillin에 감수성이던 *Streptococcus pneumoniae*가 PBP (penicillin binding protein) 변형과 같은 내성기전을 획득하면 penicillin에 내성으로 바뀔 수 있으며 이러한 경우를 획득내성이라고 한다.

앞서 설명한 것처럼 획득내성의 두 가지 기전 중 돌연변이는 하나의 염기서열에서 변이가 생기는 point mutation과 deletion, substitution, addition과 같은 비교적 규모가 큰 형태의 염기서열 재배열을 포함하며, 대표적인 예가 topoisomerase 유전자의 돌연변이에 의한 fluoroquinolone 항균제내성이다. 다른 하나인 내성 유전자의 전달은 plasmid, transposon, bacteriophage를 통해 가능하며, 세균과 세균 사이에 전달되는 방법에는 conjugation (intercellular passage of plasmid or transposon), transduction (DNA transfer via bacteriophage), transformation (uptake of naked DNA) 등이 있다. 반코마이신 내성 장알균(vancomycin-resistant enterococci, VRE)에서 vancomycin 내성을 매개하는 *vanA* 유전자가 methicillin 내성 황색포도알균(methicillin-resistant *Staphylococcus aureus,* MRSA)에 전달되어, 2002년 미국에서 처음으로 vancomycin 내성 황색포도알균(vancomycin-resistant *S. aureus*, VRSA)이 출현한 것이 대표적인 사례이다.

또한, 미생물이 항균제내성을 일으키는 유전자를 가지고 있을 때, 항균제내성은 항상 발현될 수도 있고(constitutive), 특정 항균제와 같은 유도체(inducer)에 노출된 다음에 발현될 수도 있다(inducible).

3) 다제내성

종류가 다른 여러 항균제에 동시에 내성을 보이는 경우가 많은데, 대개 서로 다른 계열의 3가지 항생제에 내성인 경우 다제내성균으로 분류한다. 이는 서로 연관이 없는 다수의 내성 유전자를 획득하거나, 하나의 유전자에 돌연변이가 생겨 서로 관련이 없는 다수의 항균제에 내성을 나타내는 기전을 통해서

가능하다. 전자의 경우는 순차적으로 다수의 내성 유전자를 획득한 후 환경에서 항균제 압력에 의해 선택되며, 후자의 경우는 한 번의 과정으로 다제내성(multi-drug resistance, MDR)이 생길 수 있다.

메티실린 내성 황색포도알균, 반코마이신 내성 장알균, 카바페넴 내성 장내세균(carbapenem-resistant Enterobacteriales, CRE)은 명칭만으로는 마치 한 가지 항생제에 내성을 가지고 있는 것처럼 생각할 수 있지만, 실제로는 여러 항생제에 내성인 대표적인 다제내성균이다.

3 항생제내성의 발생과 전파

항생제에 내성인 세균이 출현하고 증가하는 것은 항균제 사용에 따른 것임을 부인할 수는 없다. 1940년에 penicillin이 도입된 지 1년도 지나지 않아 penicillin 내성이 나타났으며, 1961년에는 MRSA가 출현하였고 1980년대 들어 본격적으로 의료기관 내에서 주류를 차지하게 되었다. 1970년대 중후반에 3세대 cephalosporin이 시장에 나왔지만 1980년대 초반에 이를 분해하고 내성을 매개하는 extended spectrum beta-lactamase (ESBL, 광범위 베타락탐분해효소) 유전자가 등장했다. 이처럼 항균제를 투여하면 항균제에 감수성이 있는 세균들은 죽고 내성균만 살아남는 현상(selection)이 생긴다. 이러한 현상은 지역사회보다는 항균제 투여가 많은 병원에서 주로 발생하며, 내성균에 감염되거나 보균하는 환자 혹은 의료진이 전파원으로 작용한다. 내성균은 주로 의료진의 손을 통해 전파되지만, 내성균에 오염된 환경이나 의료기구 등도 전파원이 될 수 있다. 하지만, 비단 의료기관뿐만 아니라 지역사회에서도 다양한 목적으로 항생제를 사용하고 있기 때문에 항생제내성의 위험이 병원 바깥에서도 문제가 된다. 지역사회 폐렴의 흔한 원인균인 폐렴사슬알균의 페니실린 내성, 성매개 감염병의 주요 원인인 임질균의 항생제내성, 식품이나 물을 매개로 감염되는 장내세균(살모넬라 등)의 항생제내성 등이 의료기관이 아닌 지역사회에서도 문제가 되는 대표적인 사례들이다.

또한, 의료기관이나 지역사회에서 사용하는 항생제는 인체 내에서 대사된 후 외부로 배설되고, 이는 하수를 통해 하천이나 강물 등 주위 환경으로 퍼져 나간다. 앞서 언급한 것처럼 항생물질이나 내성은 이전부터 자연에 존재하였고, 위와 같이 주위 환경으로 퍼져나간 항생제들 때문에 다양한 내성 유전자를 갖는 미생물들이 의료기관 밖에서, 그리고 인체가 아닌 환경에서 검출되고 있다. 최근 토양의 내성 유전체 연구결과에 따르면 토양에서 발견되는 항생제내성 유전자의 대부분이 병원이나 동물에서 분리되는 병원체에서 발견되는 동종 유전자와 동일한 것으로 밝혀지기도 했다.

항생제는 동물에서도 성장촉진제 등 다양한 목적으로 사용되고 있고, 매년 전세계에서 엄청난 양의

항생제가 축산업에 소비된다는 보고도 있다. 중국에서는 가축사료로 colistin을 사용하면서 plasmid가 매개하는 colistin 내성유전자를 가진 내성균이 가축에서 보고되기도 하였고, 유럽에서는 glycopeptide 계열의 항생제(avoparcin)를 사료에 섞어 사용하면서 vancomycin 내성균이 출현한 것이 대표적인 사례이기도 하다. 미국에서는 2002년에 사용한 항생물질의 사용량 중 약 14% 정도가 인간의 질병 치료를 위한 것이었고 그 외 83% 정도는 가축의 질병 치료나 성장 촉진제로 사용되었다는 보고가 있을 정도로, 인체 이외의 항생제 사용량이 적지 않다.

이처럼 항생제내성은 사람 혹은 의료기관에만 국한된 문제가 아니고, 환경이나 동물 등 다양한 영역과 연결되어 있기 때문에, 최근에는 사람-환경-동식물 등을 모두 포괄하는 one health(통합보건)의 개념으로 문제를 파악하고 해결하려는 시도가 중요하게 되었다(그림 8-3-1). One health는 인간의 건강, 동물의 건강, 환경의 건강 사이의 상호 의존성에 바탕을 둔 개념으로, 감염병 때문에 생기는 문제들을 해결하려면 의학-수의학-환경과학을 포함한 다양한 학문 분야의 전문가들이 협동해야 한다는 필요성을 강조하는 것이며 새로운 패러다임의 변화이다.

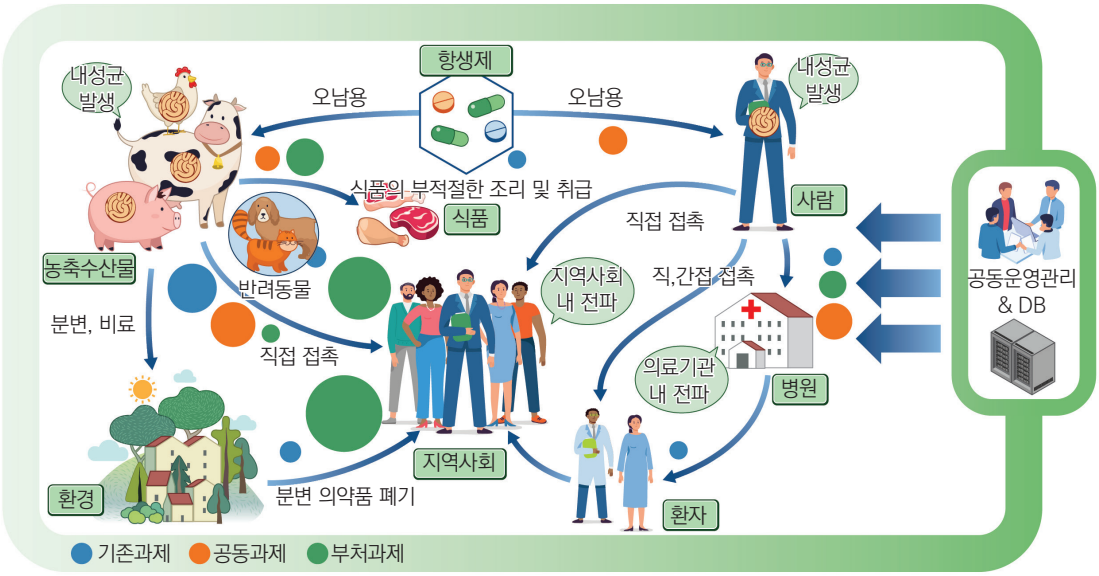


그림 8-3-1. 원헬스 기반 항생제내성 발생 및 관리 흐름도

4 항생제내성의 중요성

감염증을 치료하기 위해 항균제를 사용하는 것은 피할 수 없는 일이지만, 항균제의 오·남용은 내성균을 확산시킬 뿐만 아니라 이로 인한 부작용, 의료의 질 저하, 입원기간이나 치료기간의 증가, 입원비용(의료비)의 증가 등 여러 가지 문제를 야기한다. 항생제내성균의 발생과 유행은 치료법이 개발되지 않은 "신종 감염병" 이상의 사회경제적인 문제를 초래할 수 있어서 공중보건학적으로 매우 중요한 주제로 인식하고 있다.

2019년 미국 CDC에서는 항생제내성균 때문에 연간 280만 명이 감염되고 35,000명의 추가 사망자가 발생하며, 항생제 사용과 관련된 *Clostridioides difficile* 감염의 경우 22만여 명이 병원에서 치료를 받고 2017년에만 12,800명이 사망한 것으로 보고하였다. 물론 2013년 결과와 비교하면 항생제내성균 때문에 사망하는 질병부담이 다소 감소하기는 했지만 여전히 해결할 부분이 많다고 언급하였다.

영국에서 짐 오닐(Jim N' Oneill)이 2016년에 발간한 보고서에서는 매년 70만 명이 항생제내성 때문에 사망하고, 2050년에는 매년 전 세계적으로 천만 명의 사망자가 발생하는 것으로 예상하는 등 많은 사람들이 두려워하는 암보다 다제내성균 때문에 사망하는 환자가 더 많을 것으로 예측하였다. 이와 같은 임상적인 부담뿐만 아니라 사회경제적인 부담 역시 적지 않아서, 2050년까지 항생제내성균 때문에 매년 1~3% 정도의 국내 총생산이 감소하고, 항생제내성균 때문에 생기는 직접적인 보건의료 비용의 상승도 적지 않아서 2050년까지는 1조 달러 전후의 금액이 보건의료 비용으로 지출될 것으로 예상하였다. 이처럼 항균제내성은 의료비용의 상승을 초래하고 질병의 사망률을 증가시킬 수 있다.

최근 세계화가 가속화되면서 항균제내성의 국가간 전파도 쉬워졌기 때문에 내성문제는 어느 특정 국가에만 국한된 문제가 아니다. 2019년 자료에 따르면 전 세계에서 약 500만 명에 가까운 사람들이 다제내성균을 가지고 사망하며, 이 중 약 4분의 1에 해당되는 127만 명이 다제내성균 때문에 사망한 것으로 보고하였다.

최근 20년간 우리나라 의료기관의 임상검체에서 분리되는 주요 병원체의 항생제내성률 변화는 그림 8-3-2와 같다.

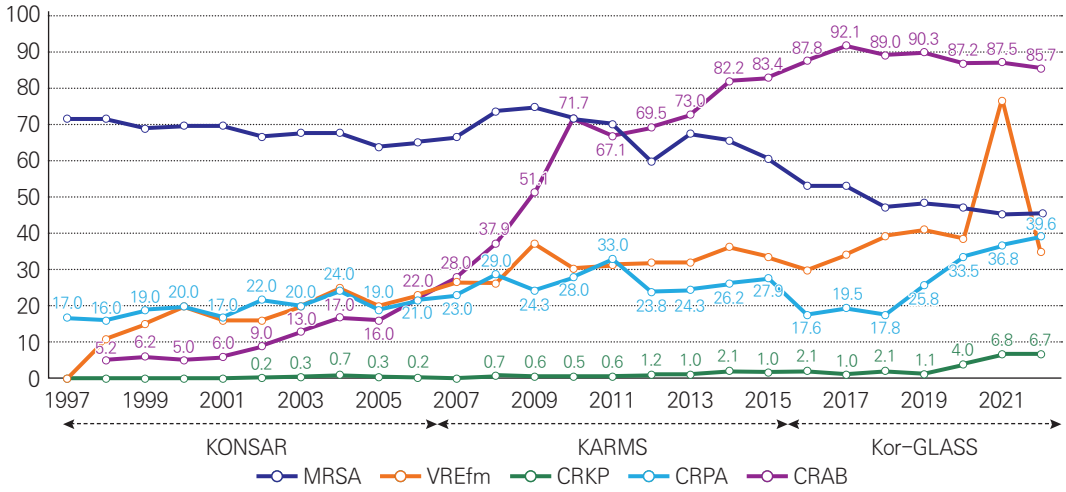


그림 8-3-2. 국내 다제내성균의 항생제내성률(1997~2022년)

우리나라에서도 법정감염병으로 신고하고 있는 6가지 다제내성균[반코마이신 내성 또는 중등도 내성 황색포도알균(VRSA/vancomycin-intermediate resistant *S. aureus*, VISA), MRSA, VRE, CRE, 다제내성 녹농균(multidrug-resistant *Pseudomonas aeruginosa*, MRPA), 다제내성 아시네토박터 바우마니균(multidrug-resistant *Acinetobacter baumannii*, MRAB)의 질병부담을 추산한 연구결과에 따르면, 매년 8,367~9,727건의 균혈증 혹은 폐렴이 생기고 이 중 3,411~3,921명이 사망하였다. 또한, 다제내성균 감염 때문에 의료비용이 추가되고 환자를 돌보기 위한 간병비가 늘어나고 조기에 사망하면서 생기는 생산성 손실을 종합하면 이와 관련된 사회적 비용은 연간 약 2,673~5,500억으로 추산하였다. 이 연구는 균혈증과 일부의 폐렴만 포함하여 추계한 것이므로 실제 항생제내성균 감염의 질병부담은 훨씬 더 심각할 것으로 예상하고 있다.

위와 같은 숫자를 가장 최근에 전세계적으로 유행했던 코로나19와 단순히 비교한다면, 전 세계에서 코로나19로 사망한 숫자가 680여만 명이었고 우리나라에서는 34,000여 명이었으므로, 매년 "조용한 판데믹"이라 여기고 있는 항생제내성의 문제가 얼마나 심각한지 가늠해 볼 수 있을 것이다.

5 항생제내성을 막기 위한 방안들

이처럼 새로운 약제가 개발되기 무섭게 바로 내성이 생기고 내성균에 감염되었을 때 치료할 수 있는 약

제가 없는 “항생제 시대의 종말”이라는 우려와 함께 내성균의 발현과 전파를 억제하기 위한 다양한 방법들이 논의되고 있다. 이러한 방법을 크게 두 가지로 나누어 보면 첫째는 항생제의 남용을 억제하고 적절한 항생제를 사용할 수 있도록 유도하는 것이고, 둘째는 병원내, 병원간, 병원과 지역사회간에 내성균이 전파되는 것을 막기 위한 방법으로 요약할 수 있다. 각각의 방법이 항생제의 적절한 사용을 유도하는 항생제 스튜어드십 프로그램(antimicrobial stewardship program, ASP)과 내성균의 전파를 차단하는 감염관리(infection prevention and control, IPC)에 해당한다. 전자(ASP)는 부적절한 항생제 사용을 줄이는 것뿐만 아니라 항생제 치료가 필요한 환자에게 최적의 항생제를 적절한 용량과 용법으로 적절한 기간 동안 투여하도록 중재하는 통합적 전략이다.

1. 항생제 스튜어드십

최근 새롭게 개발되는 항생제의 숫자가 줄어들고 내성균이 증가함에 따라 사용할 수 있는 항생제의 선택폭이 좁아지면서, 항생제를 적절하게 사용하기 위한 움직임이 점차 활발해지고 있다. ASP는 감염질환을 효과적으로 치료하는 데도 도움이 되지만, 불필요한 항생제 사용 때문에 생기는 피해를 줄여 환자를 보호하며, 항생제내성을 억제하는 중요한 전략이다. 이를 위해서는 항생제를 꼭 필요할 때만, 적절한 용량과 용법으로, 적절한 기간 동안 투여하는 것이 바람직하다. 항균범위가 광범위하고 살균력이 강하거나 혹은 반감기가 길고 조직농도가 높은 새로운 약제라고 해서 반드시 임상적으로 장점이 많고, 치료의 효율성이 우수하다는 것을 의미하지는 않는다. 또한, 경험적으로 항생제 치료를 시작할 때는 환자의 기저질환, 감염부위, 가능한 원인병원체, 그리고 지역이나 의료기관의 항생제내성률 등을 감안하여 적절한 항생제를 선택해야 하며, 원인병원체와 항생제 감수성 결과를 확인한 다음에는 항균범위가 좁은 약제로, 만약 활력징후가 안정되고 경구로 투여하는데 문제가 없다면 정맥주사 대신 경구용 항생제로 바꾸는 것을 고려해야 한다.

이와 같은 주요 전략인 ASP를 각 의료기관에 성공적으로 도입하고 정착시키기 위해 2014년 미국 CDC는 7가지 핵심요소를 제시하여 이를 수행할 수 있도록 하였다. ASP의 효과적인 실행을 위한 구조와 절차에 해당하는 핵심요소는 리더십의 책무(leadership commitment), 책임자 임명(accountability), 담당 약사 임명(pharmacy expertise), 중재의 실행(action), 추적 조사(tracking), 보고(reporting), 교육 (education)의 7가지이다. 각각의 요소를 간략하게 소개하면 다음과 같다. ① 리더십의 책무는 의료기관의 경영자들이 의료기관에서 항생제 스튜어드십을 수행하기 위해 필요한 인력, 재정, 정보자원 기술의 지원에 대한 내용이고, ② 책임자 임명은 ASP를 수행하는 전담팀을 구성하고 팀의 리더를 임명하는 것이며, ③ 담당 약사 임명은 항생제 스튜어드십을 전담할 약사를 임명하고 역할을 부여

하는 것이고, ④ 중재의 실행은 항생제를 적절하게 사용하도록 다양한 중재 방법을 의료기관 실정에 맞게 활용하는 내용이며, ⑤ 추적조사는 항생제 사용에 대한 지표 이외에도 중재의 효과를 확인할 수 있는 다양한 지표를 지속적으로 평가하고 감시하는 것이고, ⑥ 보고는 이와 같은 지표들을 경영자와 의료인을 포함하는 직원이 공유할 수 있도록 하는 것이고, ⑦ 교육은 ASP를 경영자, 의료인을 포함하는 직원과 환자들에게 교육하는 것이다. 하지만, 아직까지 우리나라에서는 ASP를 담당할 전문인력(의사, 약사 등)이 턱없이 부족하기 때문에 전문학회에서는 책임자 임명과 담당약사 임명 부분을 운영체계(operating system)로 통합하여 하나의 핵심요소로 제안하였다.

2. 감염예방과 관리

항생제내성균의 주된 전파경로는 집락 또는 감염된 환자와 직접 접촉하거나, 오염된 의료진의 손이나 주변 환경 또는 의료기구를 통한 간접접촉이다. 이를 차단하기 위한 손쉽고 비용이 적게 들면서 가장 중요한 방법은 손위생이며, 손위생을 제대로 실천할 수 있도록 시설(싱크대, 비누 및 종이타올, 격리공간 등)을 적절히 갖추어야 한다. 또한, 의료기관 내에서 전파를 막기 위해 내성균 관리체계를 마련하고, 의료종사자를 교육하며 항생제 적정사용 관리, 내성균 발생 감시활동, 선별검사, 환자 또는 병원체 보유자에 대한 격리와 접촉주의 적용, 환경관리 등이 주된 예방 전략이다.

특히, 우리나라와 같이 다인실 위주의 의료환경에서는 의료기관의 특성, 내성균의 유병률, 자원 등을 고려하여 감염관리를 집중할 다제내성균의 우선 순위를 결정해야 한다. 예를 들어, 반코마이신 장알균의 경우 그동안 우리나라 의료기관에서는 보균 혹은 감염된 환자이면 격리실 사용, 주기적 검사, 접촉주의 등 철저하게 관리했지만, 다인실 환경 때문에 점차 격리실은 부족하게 되고 이보다 더 문제가 심각하다고 평가되는 카바페넴분해효소 생성 장내세균(carbapenemase-producing Enterobacteriales, CPE)과 같은 다제내성균이 급격히 증가하면서 이전처럼 VRE 격리를 더 이상 지속하지 않는 곳이 늘어나고 있다. 왜냐하면, VRE의 병독성은 심각하지 않고, 과거와 달리 VRE 감염증을 치료할 항생제가 있으며, 내성을 매개하는 유전자가 MRSA에 전달되어 VRSA가 발생하는 사례도 급격하게 증가하지 않았지만, CPE의 경우 치료할 수 있는 항생제는 매우 제한적이고 병독성이 상당하며 앞서 언급한 것처럼 임상적-사회경제적 부담이 적지 않기 때문에 의료기관 내에서 전파를 차단하는 것이 훨씬 중요하다고 판단하기 때문이다.

3. 원헬스 접근

사람에게 중요하게 사용하는 항생제가 가축 등 동물에서도 사용되고, 항생제 오·남용으로 발생하는 항생제내성균은 사람, 동물, 환경 간에 다양한 경로를 통해 발생·전파 가능하므로 다양한 영역의 포괄적 관리가 필요하다. WHO 등의 국제기구들도 사람, 동식물, 환경의 건강이 불가분의 관계라는 'One Health' 기치 아래 항생제내성 해결을 위해 함께 노력하고 있다.

앞서 언급한 것처럼 항생제내성은 모든 사람에게 영향을 미치고, 보건-축산-수의학-농업 등 다양한 산업이 관련되어 있으므로 사회 전체의 참여가 필요하다. 또한, 비인체 분야에서도 항생제 적정 사용을 위한 ASP의 도입과 함께 농축수산 분야에서 항생제 사용을 제한하는 방안도 모색해야 하며, 비인체 분야의 항생제 사용량을 확인하고 추적관리할 수 있는 시스템을 구축해야 한다. 아울러 내성균의 발생이나 전파를 파악할 수 있는 통합된 감시체계의 운영도 필요하다. 우리나라에서도 이와 같은 목적으로 one-health 개념의 대응체계를 마련하고 포털시스템(https://nih.go.kr/nohas/common/main.do)를 운영하고 있지만, 외국의 사례와 비교한다면 앞으로 개선해야 할 숙제가 산적해 있다.

이상과 같이 항생제내성은 다양한 원인에 의해 점차 증가하고 있다. 하지만, 관련된 모든 구성원들이 적극적, 자발적으로 참여하거나 협조하지 않고서는 이를 막을 수 없으므로 사태의 심각성을 인식하여 항생제 오·남용을 줄이고 내성균의 확산을 막는 데 혼신의 노력을 기울여야 한다.

참고 문헌

1. 대한감염학회. 감염학. 3판. 파주: 군자출판사; 2024.
2. 대한감염학회. 항생제의 길잡이. 4판. 파주: 군자출판사; 2016.
3. 대한의료관련감염관리학회. 의료관련감염관리. 파주: 군자출판사; 2023.
4. 어영, 김효열, 윤갑준. 항균제와 항균제감수성 검사. 파주: 한국학술정보(주); 2007.
5. 유진홍. 감염강의 42강. 파주: 군자출판사; 2022.
6. 유진홍. 내 곁의 적. 의료관련감염. 파주: 군자출판사; 2020.
7. 유진홍. 항생제 열전. 파주: 군자출판사; 2019.
8. 장철훈, 김영아, 김현수 외. 원헬스: 사람•동물•환경. 서울: 범문에듀케이션; 2020.
9. Antibiotic Resistance Threats in the United States [Internet]. Centers for Disease Control and Prevention. 2019 [Cited 2025 Jan 3]. Available from: https://www.cdc.gov/antimicrobial-resistance/data-research/threats/index.html.
10. Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022;399:629-55.
11. Chang E, Kim HB. It Is Time to Address the Isolation Policy for Patients Colonized with Vancomycin-resistant En-

terococci. Korean J healthc assoc Infect Control Prev. 2021;26:16-23.

12. Cheong HS, Park KH, Kim B, et al. Developing Core Elements and Checklist Items for Implementing Antimicrobial Stewardship Programs in Korean General Hospitals: A Modified Delphi Survey. Infect Chemother. 2023;55:59-68.
13. Jim O'Neill. Tackling drug-resistant infections globally: final report and recommendations [Internet]. Review on Antimicrobial Resistance [Cited 2025 Jan 3]. Available from: https://amr-review.org/sites/default/files/Tackling%20drug-resistant%20infections%20-%20An%20overview%20of%20our%20work_IncHealth_LR_NO%20CROPS.pdf.
14. Kim CJ, Song KH, Choi NK, et al. Socioeconomic burden of pneumonia due to multidrug-resistant Acinetobacter baumannii and Pseudomonas aeruginosa in Korea. Sci Rep. 2022;12:13934.
15. Kim D & Jeong SH. Current status of multidrug-resistant bacteria. J Korean Med Assoc. 2022;65:468-77.
16. Park SH. Management of multi-drug resistant organisms in healthcare settings. Journal of Korean Medical Association. 2018;61:26-35.
17. Song KH, Kim CJ, Choi NK, et al. Clinical and economic burden of bacteremia due to multidrug-resistant organisms in Korea: a prospective case control study. J Glob Antimicrob Resist. 2022;31:379-85.
18. The core elements of hospital antibiotic stewardship programs [Internet]. Centers for Disease Control and Prevention [Cited 2025 Dec 5]. Available from: https://www.cdc.gov/antibiotic-use/healthcare/pdfs/hospital-core-elements-H.pdf.
19. Yoon YK, Kwon KT, Jeong SJ, et al. Guidelines on Implementing Antimicrobial Stewardship Programs in Korea. Infect Chemother. 2021;53:617-59.

제4절 기후변화와 감염병

단원 목표

기후가 감염병 전파와 유행에 미치는 영향을 설명하고, 기후변화가 유발할 수 있는 감염병 역학의 변화를 설명할 수 있다.

1 기후변화와 감염병

1. 기후변화

기후변화로 대표되는 지구환경의 변화는 21세기 인류가 맞이한 가장 큰 위협요인이다. 세계적인 산업화와 인구 증가로 대기와 해양에 축적된 온실가스는 지구환경의 비가역적인 변화를 초래하고 있다. 기후변화는 일반적으로 평균기온의 상승으로 대표되지만 기온 변동의 큰 폭의 확대와 지역 수준의 다른 요인과의 결합하여 예측하기 힘든 기상재앙과 기후의 변동으로 나타난다. 21세기 들어 기후변화의 양상은 누적된 온실가스 축적과 더불어 점차 가속화하고 있으며 21세기 중반까지 2℃ 이내 상승이라는 국제적인 목표를 달성하기 힘들 것으로 예측되어 지구환경의 비가역적인 변화가 예상된다. 이로 인해 기상이변의 빈도와 강도의 급격한 증가, 해수면 상승, 바닷물 온도 증가, 내수면 환경의 급격한 변화, 생태계 파괴 등의 재앙적 수준의 변화가 예상된다. 이는 인간의 삶을 결정하는 기저 조건의 변화를 의미하며 질병과 인간의 관계에 있어 적용되던 기존의 관계가 예측할 수 있는 범위를 크게 벗어날 수 있음을 의미한다.

2. 기후와 감염병

감염병은 기후를 비롯한 환경요인에 크게 영향을 받는다. 기후는 병원체, 숙주의 면역성과 행태, 매개체, 전파조건 등 감염병의 증식과 전파에 관련된 여러 요인에 영향을 미친다. 일반적으로 기온이 올라가면 병원체의 증식과 매개체인 절지동물의 성장과 번식에 더 유리하게 작용한다. 그러나 감염병의 전파에는 인구집단의 인구학적, 행태적 특성과 건조(建造)환경, 보건의료체계 등의 요인이 복합적으로 작용하므로 다른 요인이 고정된 이론적인 상태가 아닌 일반적인 조건에서 감염병의 행태를 기상요인만으로 예측하는 것은 위험할 수 있다, 따라서 특정 기후요인의 변동에 따른 감염병의 변동 정도는 이러한 요인들을 충분히 고려하여 해석하여야 한다(그림 8-4-1). 기상요인과 감염병 발생간의 연관성을 평가하기 위해서는 기상요인이 감염병 전파 어느 단계에 영향을 미치는지에 대한 심층적 이해가 필요하다.

기후변화로 인한 기온의 상승은 열대 및 아열대인 저위도 지역에서 유행하던 질환의 분포가 고위도 지역이나 고산지역으로 확장하는 양상을 보일 수 있다. 그러나 한편으로 저위도 지역의 해당 질병의 분포 또한 최적의 기상 조건을 초과하면 감소추세로 들어갈 수 있으므로 이러한 경향을 일률적으로 적용할 수 없는 경우가 많다.

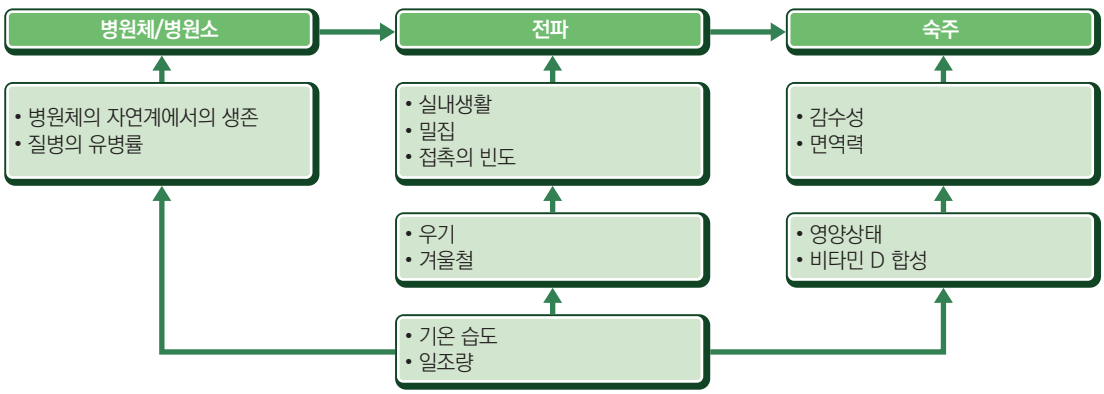


그림 8-4-1. 기후요인이 감염병의 전파에 미치는 영향

1) 계절성

감염병 중 적지 않은 경우 명백한 계절성을 띠는 경우가 많다. 특정 감염병이 특정 지역에서 계절적인 주기성을 보이는 요인에는 기후의 계절적 변화에 따른 병원체나 매개체의 변화가 일차적인 요인일 수 있으나 이외에도 인구이동 및 유입과 같은 사회경제적 요인, 실내인구 밀집, 실외활동 부족으로 인한 비타민 D 합성 부족에 의한 면역 저하 등 다양한 요인이 작용할 수 있다. 계절성을 보이는 것은 기

후 관련성의 단서가 될 수 있으나 특정 지역에서 특정 인구집단의 특정 질환에 대한 감염 관련 요인을 포괄적으로 검토해야 한다. 이를 무시하고 일률적으로 "기온 1℃ 상승하면 특정 감염병의 발생률이 몇 퍼센트 증가한다"와 같은 접근은 적절하지 않을 수 있다.

2) 지역성

기후요인에 따른 감염병의 발생 양상은 광역을 기준으로 비교적 뚜렷한 양상을 보일 수 있으나 대상 지역의 범위가 좁아질수록 기후요인 이외 다른 지리적, 사회경제적, 인구학적 요인 등이 복합적으로 작용할 수 있다. 따라서 특정 지역을 대상으로 할 경우 기상요인을 기계적으로 대입하여 감염병의 양상을 해석하는 것은 무리가 따를 수 있다.

3) 비선형성

모든 생물은 최적의 생존과 번식에 필요한 기후에 적응해 있다. 이에 따라 기온이나 강수량 등의 기상요인과 특정 생물의 번식조건은 선형적이지 않고 일정 수준 이상에서만 증식하거나(역치) 뒤집힌 U자형의 최적 온도를 가지는 경우가 많다. 기상요인과 특정 질병 혹은 생물 간의 선형적인 관계는 일정한 범위 안에서만 성립한다는 점을 염두에 두어야 한다. 이는 예년의 수준에 비하여 비교적 작은 기상조건의 변화에도 병원체와 매개체, 전파 관련 요인, 숙주 요인 등의 복합적인 변화로 이전에 없거나 드물던 감염병의 폭발적 발생이 일어나는 현상을 설명하는 중요한 요인이다.

2 매개체 감염병

매개체 감염은 사람이나 동물의 질병이 매개체를 통해 전파되는 질환으로 매개체의 생존과 번식 수준에 따라 질병의 전파 수준이 크게 좌우된다. 기후조건의 변화는 생태계의 변화에 영향을 미치며 매개체의 증식과 사람과의 접촉에도 크게 영향을 미친다. 따라서 기후요인과 매개체의 연관관계를 파악하는 것이 매개체 감염병의 유행 수준을 파악하는 데 있어 매우 중요하다.

1. 말라리아

말라리아는 오랫동안 인류의 역사와 같이해 온 질환으로 인류에게 감염을 일으키는 말라리아원충은 최

소 5종류 이상이 알려져 있다. 세계보건기구는 2022년 세계적으로 2.49억 명이 발생하여 60.8만 명이 사망하였다고 보고하였는데 이는 2000년에 비하여 현저히 줄어든 결과이지만 아직도 매개체감염병 중 으뜸을 차지하는 질환이다. 우리나라에서는 일제강점기와 한국전쟁 시기를 거치면서 큰 규모로 발생하였다가 1970년대 이후 도시화와 산업구조의 변화로 소멸하였으나 1993년 서부전선에서 근무하는 군인의 발생을 기점으로 2000년에선 연간 4,000여 명에 이를 정도로 발생하였다가 2010년 이후에는 많이 감소하였으나 여전히 연간 400명 수준으로 발생하고 있다.

기후변화에 따라 열대 및 아열대 지역의 말라리아 발생이 온난화와 함께 고위도 지역과 저위도지역의 고산지역으로 확대되는 양상을 보인다. 그러나 세계적인 말라리아 퇴치 활동이 활발하게 진행됨으로 인하여 말라리아 발생의 장기적인 양상은 방역 활동 수준을 비롯한 사회경제적 요인의 영향을 크게 받으므로 이는 일률적으로 적용할 수 없다.

국내 말라리아는 일부 해외 유입 환자를 제외하고는 모두 삼일열 원충(*Plasmodium vivax*)에 의한 것이며 국내의 주요 매개체는 중국얼룩날개모기(*Anopheles sinensis*)로 전국적으로 분포한다. 이 모기는 활동 영역이 매우 넓어 유행지역에서 감염된 모기가 일일 수백 미터에서 수 킬로미터에 달하는 거리를 이동하여 다른 사람 숙주에게 말라리아를 직접 전파할 수 있다. 국내 말라리아는 역학적으로 다음과 같은 특성을 보인다.

① 환자 발생은 모기가 활동하는 여름 시기에 집중되지만 연간 유행 시기의 초반인 8월 이전의 발생은 주로 전년도 감염 환자가 장기잠복기(long incubation period)를 거쳐 발생하는 경우가 대부분이어서 감염 시기와 발병시기의 차이가 크다.
② 환자 발생은 지역적으로 군사분계선에 접한 서부지역에 집중되어 있으며 발생률은 군사분계선에서의 거리에 반비례한다.
③ 국내 기후요인은 연간 말라리아 발생 양상에 영향을 미치나 전체 환자 발생은 접경지역 북측의 환자 발생 양상에 따라 주로 좌우된다.

국내 연구에 의하면 말라리아 환자 발생과 기후요인 중 강수량은 수주~수개월의 시간 간격(lag)을 두고 환자 발생에 영향을 미치는 것을 확인할 수 있다(그림 8-4-2). 국내 말라리아 매개 모기의 증식은 기온과 강수량에 따라 영향을 받는데 중국얼룩날개모기의 개체수는 26℃에서 가장 많다(그림 8-4-3).

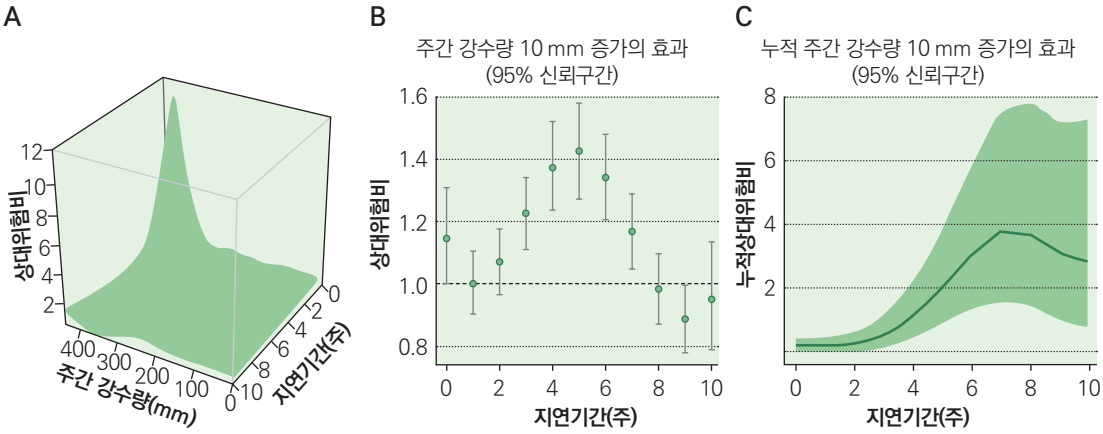


그림 8-4-2. 주간 강수량과 삼일열 말라리아 발생위험도.

(A) 주간 기온을 보정한 주간 강수량과 삼일열말라리아 발생위험도의 3-D 그래프. (B) 주간 강수량의 지연효과와 95% 신뢰구간. (C) 주간 강수량 10 mm 증가에 따른 누적 상대위험도와 95% 신뢰구간의 지연효과

[출처: Kim YM et al. 2012]

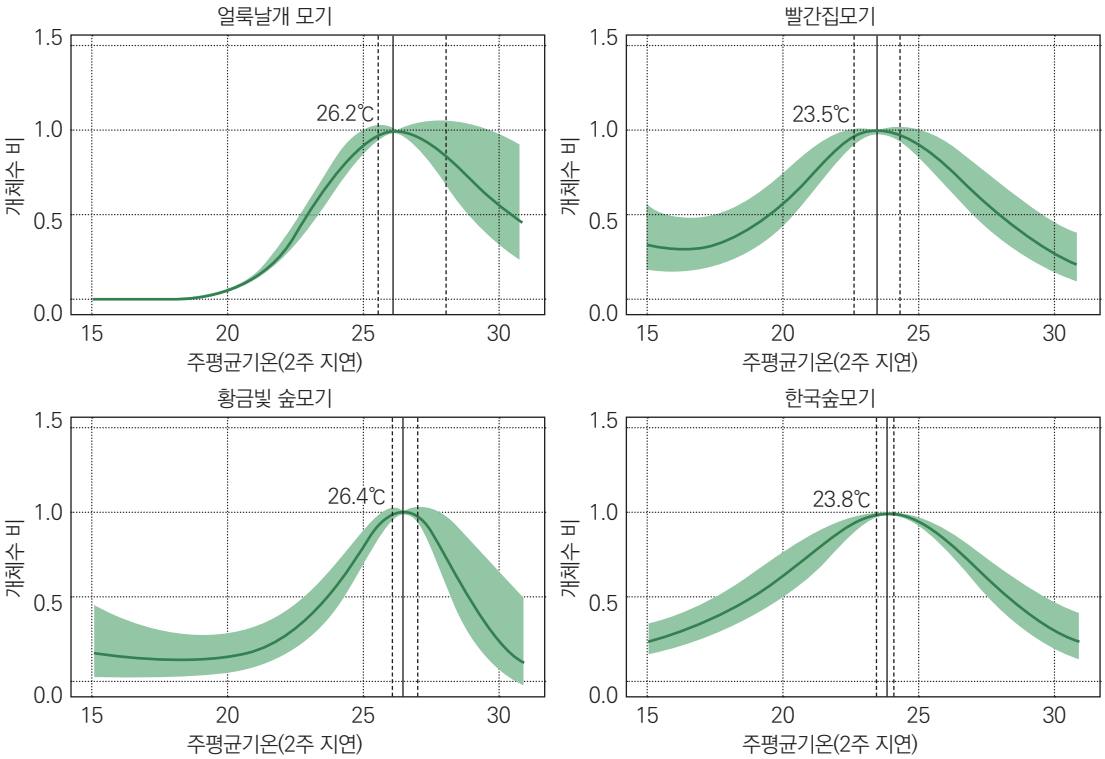


그림 8-4-3. 인천지역의 주간 기온에 따른 모기의 지역별, 종별 포획개체수

[출처: Lim AY et al. 2021]

2. 뎅기열

뎅기열은 플라비바이러스속에 속하는 RNA 바이러스인 뎅기바이러스(DENV)에 의해 생기는 질환으로 모기에 의해 전파된다. 뎅기열은 크게 뎅기열과 뎅기출혈열과 뎅기뇌염 등의 합병증이 동반된 중증 뎅기열(severe dengue)로 구분할 수 있다. DENV는 4가지 혈청형이 있으며 일개 혈청형에 감염된 이후 영구면역이 형성되므로 유행지역에서 혈청형의 분포는 집단면역 수준에 따라 달리 나타난다. 뎅기열에 대한 백신은 현재 두 가지의 백신이 승인받았으나 유행지역이 아닌 한국인에 대한 적용은 아직 제한되어 있다.

뎅기열은 1950년대 이후 지속해서 발생이 증가하고 있으며 세계적으로 연간 1~4억 명이 뎅기열에 걸리고 있으며 주로 열대 및 아열대 지역에 한정되어 있으나 해외여행 등을 통하여 온대지역으로 유입되고 있고 온대지역에서도 2014년 일본 도쿄 요요기공원에서와 같은 일부 소규모 집단발병이 보고되고 있다. 국내에서는 뎅기열의 발생은 전적으로 해외 유입에 국한되어 있으며 2000년대 이후 보고된 이후 해외여행자의 증가와 더불어 환자 수가 증가하고 있다. 국내에서의 첫 유행은 1940년대 남부지역에서 발생이 의심되고 있지만 공식적으로는 확인되지 않았다.

뎅기열을 전파하는 매개체는 이집트숲모기(*Aedes aegypti*)와 흰줄숲모기(*Aedes albopictus*)를 통하여 주로 전파되고 있으며 흰줄숲모기는 주로 열대지역에 서식하는 이집트숲모기에 비하여 전파 능력이 떨어지지만, 국내에 광범위하게 서식하고 있어 국내에서의 자생적인 발생 가능성은 남아 있어 대비가 필요하다.

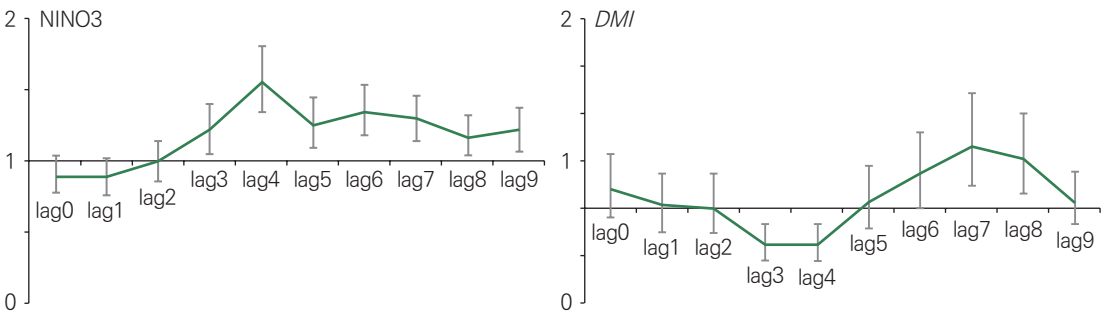


그림 8-4-4. 국내 주간 여행지별 해외 유입 뎅기열 환자 발생률과 거시 기상지표
NINO3(엘니뇨 남방 진동 지표)는 4주(좌), DMI(인도양 진동 지표)는 7주 앞서 환자 발생을 예측할 수 있음
[출처: 성균관대학교]

국내 유입되는 뎅기열은 연간 200명 수준으로 주로 해외여행자 수가 많은 여름과 겨울철에 집중되

어 있으나 여행자 수를 보정한 발생률은 뎅기열의 주요 발생지역인 동남아시아 및 남아시아 지역의 전체 환자 발생 수를 민감하게 반영하고 있다. 이들 유행지역의 환자 발생 수준은 지역단위로는 이전 유행에 따른 집단면역 수준에 따라 차이를 보이나 지역 전체의 발생률은 중장기적인 기후요인의 변화를 민감하게 반영하고 있다. 이에 따라 국내의 해외 유입 뎅기열의 환자 발생은 엘니뇨 남방진동(El Niño Southern Oscillation, ENSO)과 인도양 쌍극자(Indian Ocean Dipole, IOD), 남아시아 몬순 등의 거시기상지표에 따라 비교적 정확하게 그 발생을 예측할 수 있다(그림 8-4-4).

뎅기열의 국내 발생은 2014년 도쿄 중심부 요요기 공원에서 발생한 집단발병 사례가 국내의 여건과도 크게 다르지 않으므로 여름철 대도시를 중심으로 집단발병의 가능성이 남아 있어 모기모니터링과 환자감시체계를 통한 철저한 대비가 필요하다. 장기적인 기후변화를 감안할 때 국내에서도 이번 세기 후반부에는 제주를 비롯한 남부 해안 지역은 유행잠재력이 높은 이집트숲모기의 서식이 가능한 조건이 될 것으로 예상되므로 뎅기열의 국내 정착에 대한 대비가 필요하다. 또한 뎅기열은 치쿤구니아열, 지카바이러스 감염증 등의 아르보바이러스 감염증과 감염경로와 양상이 비슷하므로 뎅기열에 대한 감시체계 및 비상대응대책은 이들 질환을 포함하여 수립할 필요가 있다.

3. 쯔쯔가무시증

쯔쯔가무시증은 1980년대 국내에서 보고된 이후 국내 매개체 감염병 중 가장 많은 환자가 발생하는 질환이다. 주로 늦가을 농촌지역에서 집중적으로 발생하는데 해마다 계속 증가하는 추세였으나 2019년 이후 전체 환자 수가 줄어들었다가 다시 늘어나는 추세이다. 쯔쯔가무시증은 *Orientia tsutsugamushi* 균에 감염된 털진드기 유충에 의해 전파되는 질환으로 국내에는 51종의 털진드기과 진드기가 있으나 이 중 8종이 쯔쯔가무시증을 매개하며 특히 활순털진드기(*Leptotrombidium scutellare*)와 대잎털진드기(*Leptotrombidium pallidum*)의 2개 종이 가장 중요하다. 전국적으로 고루 분포되고 봄과 가을철에 발생하는 대잎털진드기에 비하여 활순털진드기는 주로 호남, 경남, 제주 남부지역에서 우점종으로 주로 가을철에 발생하는 지리적, 시간적 차이를 보인다(그림 8-4-5). 털진드기는 유충 시기에 흡혈하므로 사람에 감염도 유충이 활동하는 시기와 연관되어 있다. 유충이 활동하는 시기인 여름철의 기상요인이 그해 가을의 환자 발생과 연관되어 있다. 한편 농촌인구의 노령화와 더불어 감염 취약인구의 증가 역시 장기적으로 국내 발생 양상의 악화와 연관성이 있다. 쯔쯔가무시증은 향후에도 국내에서 가장 중요한 진드기 매개 질환일 것으로 예상된다.

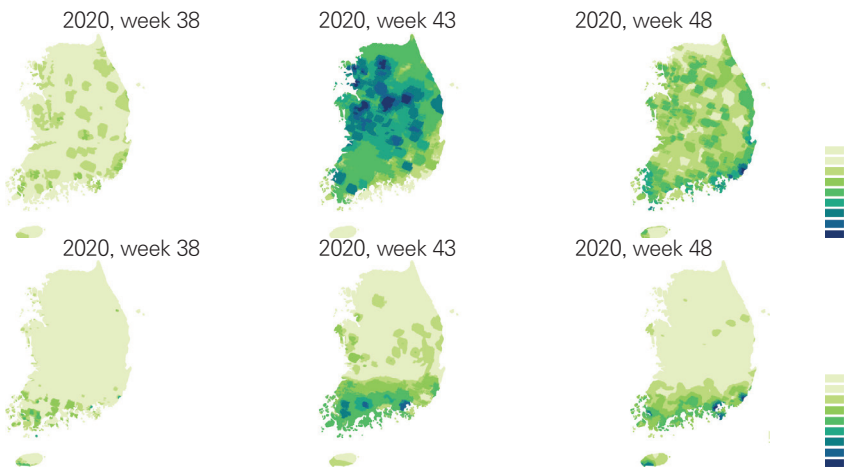


그림 8-4-5. 국내 털진드기의 시기별 지리적 분포
(상) 대잎털진드기, (하) 활순털진드기
[출처: 질병관리본부, 2019]

한편 소지역 단위에서 볼 때 진드기 매개 질환의 환자 분포는 불균등 분포를 하는 경우가 많고 지역적인 집락이 관찰되는 경우가 많다. 특정 지역에서 진드기가 고밀도로 분포하는 곳을 진드기 군집지역 혹은 진드기 집중서식지(mite island)라고 하는데 지역 내 집락은 이와 같은 진드기 군집지역을 찾아내는 것이 예방과 관리를 위해 매우 중요하다.

3 수인성 식품매개 감염병

수인성 식품매개 감염병은 소화기에서 증식한 병원체가 대변을 통해 배출된 다음 식품이나 물과 같은 매개체를 거치거나 오염된 손이나 식기 등을 통해 직접 전파되는 질환이다. 외부로 배출된 병원체는 식품에서는 박테리아의 증식이 가능할 수 있지만, 물에서는 증식할 수 없어 시간 경과에 따라 소멸하므로 병원체의 장기간 생존능력과 매우 낮은 농도 병원체의 침입으로도 숙주에서 감염이 성립될 수 있는 병원체의 생존능력과 감염력이 중요하다. 일반적으로 외부 기온이 높으면 외부 환경에서의 생존이 박테리아에는 유리하나 바이러스의 경우에는 그렇지 못하다. 이와는 별도로 계절에 따른 숙주인 인간의 행동양식과 사회경제적 환경의 변화는 소화기감염증 발생의 주요한 요인이 될 수 있다.

우리나라의 소화기감염증의 발생위험과 평균기온 간의 관계를 보면 0℃와 20℃를 변곡점으로 하

여 N자형의 양상을 보인다(그림 8-4-6). 이와 같은 분포는 낮은 기온에서는 주로 바이러스성 감염이, 높은 기온에서는 박테리아성 감염이 주로 발생하는 두 가지의 양상이 혼합된 결과로 볼 수 있다(그림 8-4-7).

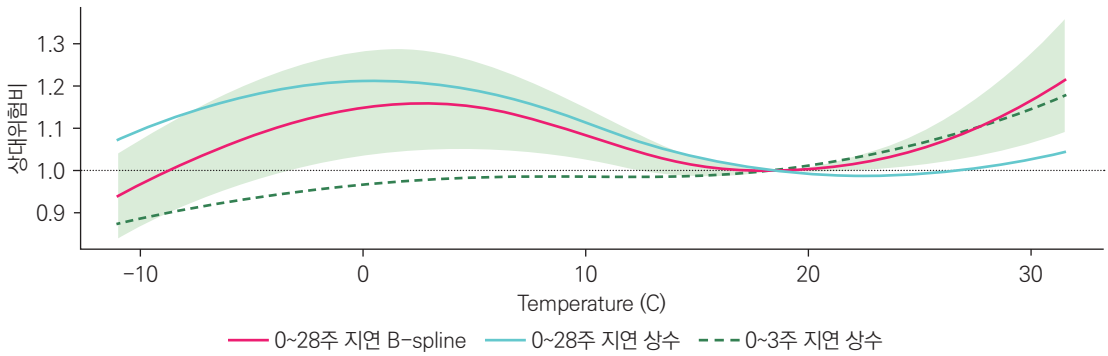


그림 8-4-6. 기온과 장 감염질환 발생의 상대위험도
[출처: 질병관리청, 2022]

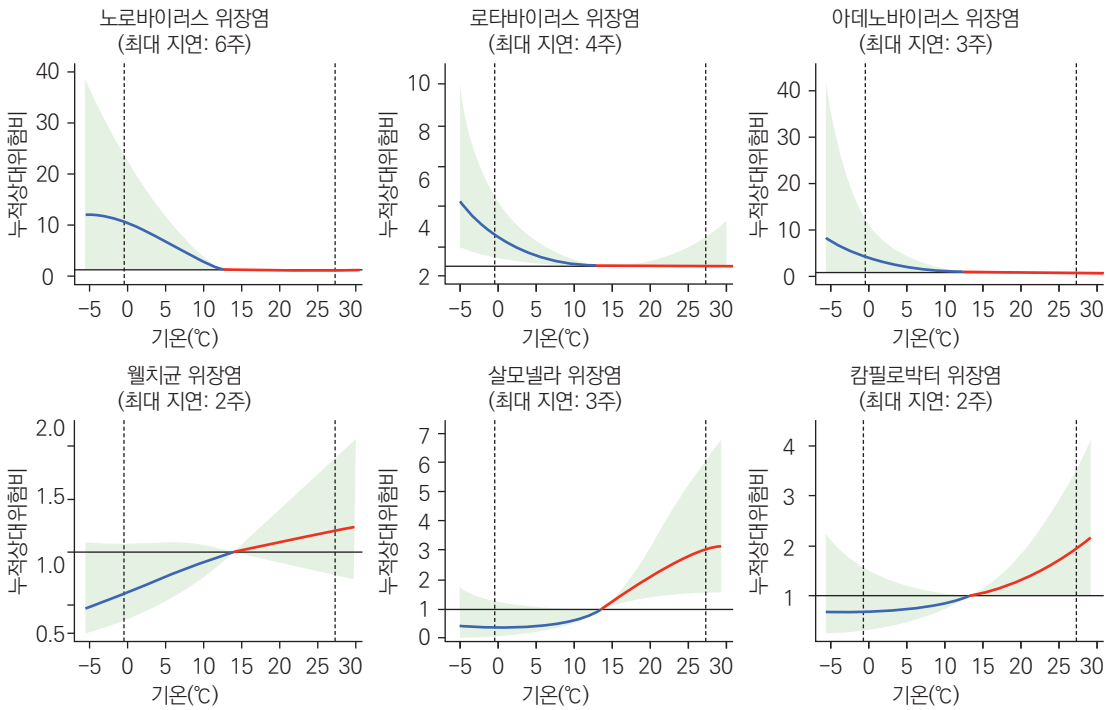


그림 8-4-7. 국내 주요한 위장관감염 병원체의 월평균기온에 따른 누적 발생위험도, 2015~2019
[출처: Sung JS et al., 2022]

강수량과 소화기 감염병의 분포는 지역의 특성에 따라 매우 다르게 나타나는데, 정수된 물에 대부분의 경우 접근이 가능한 우리나라의 경우 재난 상황과 같은 특수한 경우를 제외하고는 강수량과 감염병 간의 관계를 확인하기 힘들다. 그러나 안전한 물 공급이 원활하지 않은 지역에서는 강수량은 소화기 감염병과 매우 밀접한 연관성을 보일 수 있다. 강수량의 계절적 차이가 큰 지역에서는 강수량과 소화기 감염병의 발생이 U자형의 분포를 하는 경우를 보는 일도 있는데 이는 가뭄 때에는 생활용수의 부족으로 인한 위생관리의 미비로 집적전파가 일어나는 반면 비가 많이 오면 오염 물질의 이동으로 상수원 오염에 취약해진 결과로 감염이 증가하는 것으로 해석할 수 있다.

4 호흡기 감염증

호흡기 감염증 발생의 계절적 요인은 인구의 밀집도와 실내생활 시간, 인구집단의 면역수준 등의 요인이 복합적으로 작용한다. 일반적으로 실내 생활시간이 늘어나는 겨울철에는 호흡기질환 발생이 증가하는 경우가 많다. 겨울철은 또한 일조량이 적고 옥내생활 시간이 늘어남에 따라 비타민 D 합성이 감소하여 면역력의 저하로 이어질 수 있다. 이러한 이유로 일반적으로 호흡기감염증은 겨울철에 주로 발생하는 양상을 보인다. 그러나 COVID-19과 같이 감염력이 매우 강력하고 집단면역이 형성되지 않은 질환의 경우 기상요인이 감염병 발생에서 차지하는 비중은 상대적으로 낮을 수 있다. 열대 지방의 경우 소아 폐렴의 발생은 실내생활 시간이 증가하는 우기에 더 많을 수 있지만, 경제활동이 저하되고 영양상태도 나빠지는 건기에 발생률이 높아지기도 한다. 이처럼 특정 감염병의 계절성은 해당 지역 인구집단의 특성과 사회경제적 요인 등이 복합적으로 작용하여 나타나게 되므로 지역에 따라 그 요인을 분석하고 그 결과를 해석하여야 한다.

참고 문헌

1. 질병관리본부, 성균관대학교. 국내 뎅기열 토착화 예측모형 개발 연구. 2016.
2. 질병관리본부, 성균관대학교. 국내 말라리아 퇴치를 위한 중장기 전략 개발. 2017.
3. 질병관리본부, 성균관대학교. 해외유입 모기매개감염병 국내 발생 대응 시나리오 개발. 2018.
4. 질병관리청, 성균관대학교. 기후변화 매개체 전파 감염병 예측 모델 및 평가 연구. 2022.
5. 질병관리청. 제1차 기후보건영향평가 보고서. 2022.

6. Kim J, Kim JH, Cheong HK, et al. Effect of Climate Factors on the Childhood Pneumonia in Papua New Guinea: A Time-Series Analysis. Int J Environ Res Public Health. 2016;13(2):213.
7. Kim YM, Park JW, Cheong HK. Estimated effect of climatic variables on the transmission of Plasmodium vivax malaria in the Republic of Korea. Environ Health Perspect. 2012;120(9):1314-9.
8. Lim AY, Cheong HK, Chung Y, et al. Mosquito abundance in relation to extremely high temperatures in urban and rural areas of Incheon Metropolitan City, South Korea, from 2015 to 2020: an observational study. Parasit Vectors. 2021;14(1):559.
9. McIver LJ, Naiker J, Hales S, et al. Climate change and human health in Fiji: environmental epidemiology of infectious diseases and potential for climate-based early warning systems. Fiji J Public Health. 2012;1(1):7-13.
10. McMichael AJ. Globalization, climate change, and human health. New Engl J Med. 2013;368:1335-43.
11. Park JW, Cheong HK, Honda Y, et al. Time trend of malaria in relation to climate variability in Papua New Guinea. Environ Health Toxicol. 2016;21:e2016001.
12. Sung J, Cheong HK, Kwon HJ, et al. Pathogen-specific response of infectious gastroenteritis to ambient temperature: national surveillance data in the Republic of Korea, 2015-2019. J Hyg Environ Health. 2022;240:113924.

Textbook For
Korea Field Epidemiology
Training Program





9장

과학적 보고서 작성과 발표







제1절 역학논문 비평적으로 읽기
제2절 체계적문헌고찰 논문의 해석
제3절 감시분석보고서의 작성
제4절 역학조사 보고서의 작성
제5절 초록 작성 및 학술 발표
제6절 STROBE 및 ORION 가이드라인의 이해:
역학조사 결과의 논문 작성 및 연구결과의 공표

집필자 김현정, 손현진, 유석주, 천병철

제1절 역학논문 비평적으로 읽기

1 연구설계의 구분

역학연구(epidemiological research)는 질병, 건강 문제, 건강 상태의 분포 및 결정 요인에 대한 연구이다. 사람들 사이에서 발생하는 질병과 건강의 상태를 이해하고 이를 예방하거나 제어할 수 있는 방법을 모색하는 데 중점을 두며 질병의 원인, 유병률, 발생요인 등을 파악하기 위해 주로 수행된다. 역학연구에는 환경역학, 유전역학, 영양역학 등 많은 역학 하위 분야가 포함되며 이에 따라 보다 다양한 방식의 연구가 수행되고 있다.

역학논문을 비평적으로 읽기위해서는 기본적으로 연구설계(study design)의 특성을 이해하는 것이 필요하다. 연구설계는 크게 관찰연구와 실험연구로 구분할 수 있으며, 관찰연구는 연구자가 실험을 하지 않고, 자연상태에서 사람들의 특성이나 행동을 관찰하여 데이터를 수집하는 방식이다. 주로 질병의 위험 요인이나 예후를 파악하는 데 사용되며 역학연구에서 주로 사용되는 연구방법이다. 일반적인 관찰연구설계는 다음과 같다.

1. 코호트연구

일정한 시간 동안 특정 집단을 추적 관찰하며, 질병의 발생과 그 원인 사이의 관계를 분석하는 연구방법이다. 코호트연구(Cohort Study)의 시작시점에 따라 현재 시점에서 코호트를 식별하고 일정 기간을 추적하는 연구방법(전향적 코호트연구)과 과거시간의 코호트를 식별하고 현재까지 추적하는 연구방법(후향적 코호트연구)으로 구분할 수 있다. 전향적 및 후향적 설계에서 코호트연구는 원인에서 결과의 추적이라는 동일한 방향으로 진행된다.

또한 코호트연구는 드문 노출 연구에 유용하다. 예를 들어, 코호트연구는 화학 물질과 같은 직업적 노출의 건강 영향을 조사하는 데 적합한 연구설계이다. 더욱이 코호트연구는 생존자 바이어스의 위험을 줄여준다.

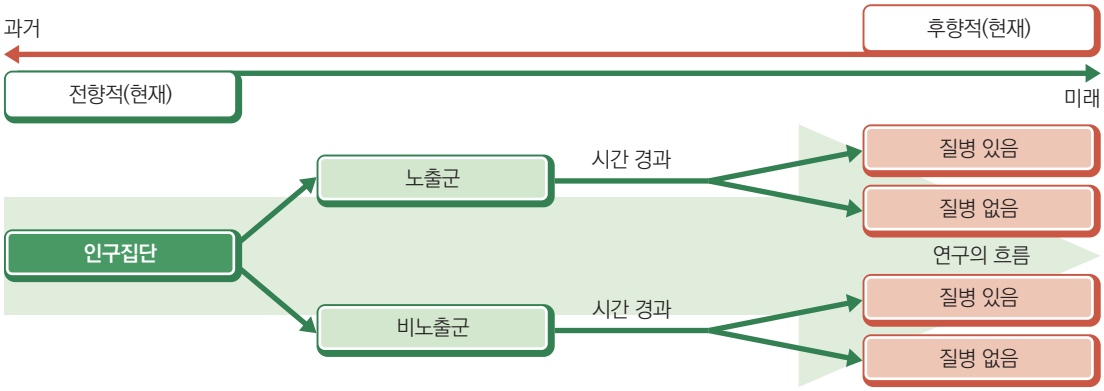


그림 9-1-1. 코호트연구 흐름과 연구시작 시점

모든 연구 설계와 마찬가지로 관심 노출을 제외하고 나머지 모든 특성이 비교군에서 동일해야 한다. 그러나 실험 연구와 달리 코호트연구에서는 연구 대상 그룹 간에 비교 가능성이 없는 경우가 많아, 이로 인해 선택 바이어스(selection bias)가 발생할 수 있다. 또한 코호트연구는 암과 같이 희귀 질병이나 발병하는 데 오랜 시간이 걸리는 질병의 경우, 비용을 포함한 여러 문제로 인해 최적의 연구 설계가 아니다. 특히 ① 추적 조사에서 탈락이 많은 경우, ② 노출여부에 따라 추적조사 탈락의 차이가 있는 경우, ③ 추적기간 동안 연구 참여자의 노출 상태가 바뀔 가능성이 있는 경우 연구결과를 왜곡시킬 가능성이 있다. 또한 후향적 코호트는 특히 연구자의 의도가 반영되지 않는 장점이 있으나 특정 노출과 건강결과 간의 관련성이 의심되는 경우 노출여부에 따른 검출바이어스(detection bias)가 발생할 가능성을 배제할 수 없는 문제점이 있다.

2. 환자-대조군연구

질병 혹은 관심결과가 있는 집단과 없는 집단을 비교하여, 질병발생과 관련된 위험요인을 분석하는 방법이다. 이 연구설계에서 연구 그룹은 결과에 따라 정의되고 연구 대상 집단은 관심 질환 혹은 결과가 발생했는지 여부에 따라 사례(환자)와 대조군으로 구분된다. 환자-대조군연구(case-control study)는 발생률을 산출할 수 없지만 사례그룹과 대조군 각각에 노출된 개인의 비율에서 파생된 승산비를 산출한다. 또한 관심 대상 집단에서 특정 결과의 발생률이 낮을 때, 환자-대조군연구의 승산비는 위험 비율의 좋은 근사치로 간주될 수 있다.

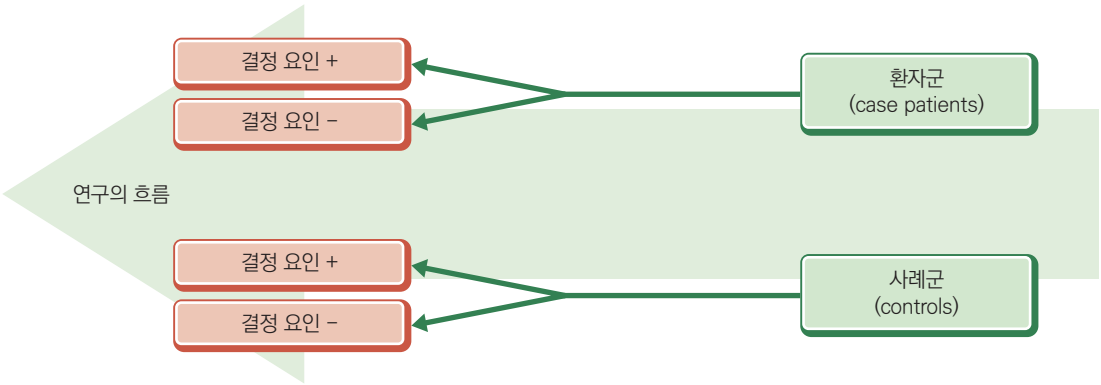


그림 9-1-2. 환자-대조군연구 흐름도

환자-대조군연구는 시간, 비용, 노력 측면에서 효율적인 연구설계이다. 일반적으로는 암과 같이 발병률이 낮은 질병이나 잠복기가 긴 질병을 조사하는 데 적합하다. 그러나 노출 빈도가 낮으면 연구자가 노출된 사람을 찾기 위해 많은 사례와 대조군을 조사해야 하기 때문에 적절하지 않을 수 있다. 환자-대조군연구는 대조군의 선택과 노출정보 획득과 관련된 주요 방법론적 문제가 발생할 가능성이 있다.

1) 주요 바이어스

① 선택바이어스는 연구(또는 분석)에 개인을 선택하는 과정에서 발생하는 문제로 결과와 노출의 관련성에 영향을 미치게 된다. 이러한 선택바이어스를 예방하기위해 관심 결과를 가진 연구대상과 관심 결과를 가지지 않은 연구대상이 동일한 특성이 있어야 하며, 주로 관심 결과를 가진 연구대상과 같은 대표성을 가진 집단에서 대조군을 선택하도록 하고 있다. 예를 들어, 환자군과 대조군은 동일한 지역의 일반 인구을 대상으로 무작위로 선택한다면 조사자의 의도와 판단을 최소화할 수 있다.

② 정보바이어스(information bias)는 노출, 공변량 또는 결과 변수 측정의 결함으로 인해 환자군과 대조군 그룹 간에 정보의 질이 달라진 것을 의미한다. 정보바이어스 중 환자-대조군연구에서 가장 많이 발생하는 유형은 회상바이어스(recall bias)로 과거 사건이나 경험에 대한 회상의 정확도나 완전성의 차이로 인한 체계적 오류를 의미한다. 예를 들어, 흑색종 위험 요인에 대한 환자-대조군연구에서 과거의 일광 노출, 일광 화상 및 일광욕실 사용에 대한 정보를 후향적으로 수집하면 흑색종을 진단받은 대상자가 이러한 요인에 대한 노출 증가를 보고할 가능성이 더 높을 수 있다. 따라서 주요 관심요인과 결과간의 관련성을 과장시킬 수 있다.

3. 단면 연구

단일시점에서 인구 집단을 대상으로 데이터를 수집하여, 질병(특정 관심 상태)의 유병률과 관련 요인 분석하기위해 수행되며 노출과 결과가 동시에 확인되는 것이 중요한 특성이다. 단면 연구(cross-sectional study)는 주로 임상 연구에서 질병의 유병률을 이해하는 데 사용되며, 유병률은 질병이 처음 발생한 시점과 관계없이 주어진 시간에 특정 질병이나 속성을 가진 인구의 비율을 말한다. 가장 대표적인 사례는 국민건강영양조사 자료를 이용한 연구이다.

결과 및 노출 변수가 동시에 측정되기 때문에 단면 연구에서 인과 관계를 도출하는 것은 어렵지만 일반적으로 빠르고 저렴하게 수행할 수 있다는 장점을 갖는다.

2 역학연구의 평가를 위한 주요 고려사항

1. 내적 타당도(Internal Validity)

연구설계, 데이터수집 과정을 통해 결과가 올바르게 도출되었는지를 의미하는 개념으로 정확성과 신뢰성을 의미한다.

2. 외적 타당도(External Validity)

연구 결과가 특정 인구 집단에만 적용되는 것이 아니라, 더 넓은 범위에서 일반화될 가능성을 의미한다. 일반적으로 연구대상자의 대표성, 환경적 요인, 시간적 요인 등이 외적타당도에 영향을 미치게 된다.

3. 바이어스(Bias)

연구 설계나 데이터 수집 과정에서 의도하지 않게 발생하는 오류로, 연구 결과의 정확성을 왜곡할 수 있다.

4. 교란변수(Confounding variable)

연구에서 분석하고자 하는 변수와 관련이 있는 다른 변수들이 존재하고 이들이 연구 결과에 영향을 미치는 경우를 의미한다. 일반적으로 교란요인은 질병에 대한 외부 위험 요인이며, 노출과 관련이 있어야 하나 노출이나 질병의 영향을 받거나 노출과 관심 질병 사이의 인과 경로에서 중간 단계(매개요인)가 되지 않는다는 특성을 갖는다. 교란 요인을 통제하기 위해 연구설계 중 매칭 혹은 제한을 하거나 사후에 통계적으로 보정하는 것을 선택할 수 있다.

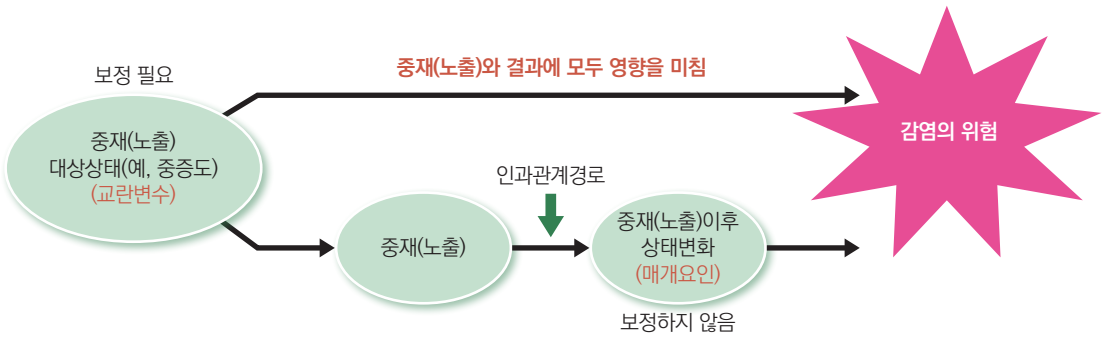


그림 9-1-3. 교란변수와 매개요인의 차이

3 연구설계별 평가도구

논문의 비평적 평가는 일반적으로 연구설계 별로 개발된 도구를 적용하여 수행하게 된다. 관찰연구의 경우 다양한 연구설계가 존재하므로 모든 관찰연구를 통합하여 평가할 수 있는 하나의 도구가 개발되어 있지 못하다. 각 연구설계 별로 타당도가 검증된 도구를 선택하여 적용할 수 있다.

관찰연구의 바이어스 위험을 평가하기 위해 개발된 도구 중 국내 연구자가 개발한 도구가 RoBANS 2(RoBANS 2: A Revised Risk of Bias Assessment Tool for Nonrandomized Studies of Interventions)이며, 이 도구는 각 평가항목 마다 연구설계에 따른 세부 평가지표를 제시한 장점을 갖는다. 기본적으로 중재를 대상으로 하나, 노출을 중재의 개념으로 접근하는 경우 사용가능하다. 코크란의 방법론 그룹에서 최근 제시한 ROBINS-E (Risk Of Bias In Non-randomized Studies of Exposures) 도구는 코호트연구를 평가하기 위해 개발된 도구이다. 이전 ROBINS-I (Risk Of Bias In Non-ran-

domized Studies of Interventions)도구와 유사하나 기존 도구는 중재를 대상으로 하였다면 노출을 대상으로 한 관찰연구를 평가하기위해 새로 개발되었다. 기본적으로 두 도구 모두 코호트, 사전사후연구, 시계열연구, 비교군 없는 연구 모두 대상으로 하며, ROBINS-E의 경우 환자-대조군연구는 추가 개발될 예정이다. ROBINS-E 도구와 RoBANS 2의 평가항목은 다음과 같다. 두 도구의 평가항목은 거의 유사하나 ROBINS-E 도구는 각 평가항목별 하위 질문으로 구성되어 알고리즘을 따라 평가하므로 좀 더 쉽게 적용할 수 있는 장점이 있다.

표 9-1-1. 두 관찰연구의 평가도구의 평가항목 및 바이어스 유형

바이어스유형	ROBINS-E	RoBANS 2
선택바이어스(selection bias)	교란 요인	교란요인
		대상집단의 비교가능성
	연구(또는 분석) 참가자 선정	대상집단 선택
	노출 이후 중재	
수행바이어스(performance bias)	노출 측정	중재/노출의 측정
선택손실바이어스(selective attrition bias)	결측된 데이터	불완전한 결과자료
검출바이어스(detection bias)	결과 측정	평가자의 맹검
		결과측정
보고바이어스(reporting bias)	보고된 결과 선택	선택적 결과 보고

1. 교란요인으로 발생한 바이어스의 위험

두 도구 모두 교란요인에 대한 평가의 경우 사전에 발생 가능한 교란요인을 목록화하여 해당 연구가 ① 중요 교란요인을 고려하였는지, ② 교란요인을 측정하기위해 선택된 변수가 신뢰성과 타당성 있는 방법으로 측정되었는지, ③ 음성대조군(negative control) 또는 기타 방법을 사용하여 미측정 교란 혹은 잔존 된 교란요인의 가능성이 배제되었는지, ④ 관련 교란요인을 모두 통제하였는지에 대한 항목을 평가하도록 되어 있다.

중요한 교란요인을 모두 고려하는 것이 필요하며 교란요인을 고려한 경우에도 교란요인을 측정하기 위해 선택된 변수 혹은 도구가 신뢰성과 타당도가 입증되었는지를 확인하는 것이 필요하다. 또한 대리지표가 사용되었다면 잔존 된 교란의 효과에 대한 고려도 필요하다.

음성대조군은 측정되지 않는 교란요인의 존재를 확인하는 하나의 방법이다. 예를 들면, 인플루엔자 백신 이후 인플루엔자로 기인한 사망과 입원에 대한 예방효과 연구에서 결과에 대한 통제로서 인플루

엔자로 기인한 사망이 아니라 사고로 기인한 사망이 백신군에서 더 적다면 이는 백신의 효과가 아닌 다른 요인이 작용한다는 것을 의미한다. 이러한 대조군을 통해 고려되지 않은 교란요인이 없는지에 대한 확인이 필요하다.

관련 교란요인의 통제는 연구설계단계에서 연구분석단계까지 다양하게 진행될 수 있다. 연구를 평가하기 위해서는 사전에 교란요인을 목록화하여 해당 연구가 중요한 교란요인에 대한 통제가 연구진행단계 중 어느 단계에서 통제되었는지 확인하는 것이 필요하다. 일반적인 통제 방법은 층화(stratification), 회귀 분석(regression), 매칭(matching), 표준화(standardization), 역확률 가중치법(inverse probability weighting, IPW) 등이다. 관찰연구를 비평적으로 평가할 때 가장 중요한 고려사항은 비교군 간의 비교가능성이다. 따라서 반드시 교란요인이 통제된 상태의 결과값을 사용하거나 해석해야하며 이러한 통제가 없는 연구 결과 인 경우 사용 여부에 대한 선택이 필요하다, ROBINS-E의 경우 교란요인에 대한 통제가 되어 있지 않는 연구는 비뚤림 위험이 매우 높음으로 향후 평가를 진행할 필요성이 있는지에 대한 판단을 하라고 제시하고 있다. 교란요인은 일반적으로 잘 알려진 기저 교란요인(Baseline confounding)이 있지만 이외에도 시간과 관련된 시간변동교란요인(Time-Varying Confounding)도 있다. ROBINS-E의 경우는 이러한 시간과 관려된 교란요인에 대해 좀더 엄격한 평가 항목을 제시하고 있어 기존의 도구와의 차이가 있다. 시간변동교란요인은 추적기간 중 노출의 상태가 변화하는 것으로 결과와 관련성이 있을 때 발생가능하다. 예를 들어 B형간염 환자에서 스타틴 복용이후 간암발생과의 관련성을 보는 연구에서 간염환자의 경우 스타틴은 간독성이 있는 것으로 알려져 있다. 따라서 스타틴을 복요하다 간수치가 높아지는 경우 스타틴을 복용중지 될 가능성이 있으며 간수치가 높은 대상자는 간암 발생 가능성이 또한 높을수 있다. 이 경우 간수치는 대상자의 추적기간 중 노출의 상태를 변화시키고 결과에도 영향을 미치는 시간변동교란요인이다.

2. 노출의 측정과 관련된 바이어스위험

노출 측정의 평가는 측정횟수와 방법에 따라 차이가 존재한다. 노출에 대한 단일 측정값이 충분할 수 있는 경우는 노출이 시간 경과에 따라 안정적이거나, 반감기가 길거나, 혹은 관심 노출이 특정 상태일 때이다. 만약 단일 측정값이 충분하지 않다면 여러 번 측정한 경우가 고려되며, 이 경우는 노출 기간과 양적 특성을 잘 조합하는 노출지표를 선택하였는지에 대한 평가가 필요하다. 노출의 기간과 양적 특성을 고려하는 것은 매우 중요하다. 예를 들어, 빠르게 변화하는 혈중 대사물질이나 대기오염과 같이 누적 노출이 관련된 경우, 특정 시점의 노출이 관심 있는 노출을 제대로 반영하지 못하게 된다. 또한 환경측정을 개별 노출로 간주하거나 노출의 영향에 대한 잘못된 가설을 가지고 있는 경우, 샘플측정이 실제

값과 차이가 있더라도 부적절한 결과를 도출할 수 있다. 따라서 관심 노출의 특성을 파악하는 것이 필요하다.

노출 측정과 결과 여부의 관련성에 따라 차별적 오류와 비차별적 오류로 구분될 수 있다. ① 비차별적 오류는 결과측정과 무관하게 발생하는 오류로 예를 들어 측정값을 정수로 반올림하는 과정에서 발생할 수도 있다. 일반적으로 무작위적 오류는 노출효과를 무효가설 방향으로 바이어스시켜서 회귀희석 바이어스(regression dilution bias)로 불리기도 한다. ② 차별적 측정오류는 노출의 측정이 건강결과 혹은 건강결과의 위험과 관련되어 결과 여부에 따라 다르게 측정되거나 분류되는 것을 의미한다. 측정 방향에 따라 연구결과를 과장 혹은 과소 추정하게 된다. 노출과 결과가 동시에 측정되는 경우인 단면연구나 결과가 발생한 이후 노출을 측정하는 환자-대조군연구에서 더 자주 발생하나 후향적 코호트연구에서도 발생할 수 있다.

노출에 대한 내용을 평가하기 위해 고려할 점은 관심있는 노출에 대한 적절한 지표인가? 노출이 오류로 측정되었거나 잘못 분류될 가능성은 없는가? 노출의 오분류 및 잘못된 측정이 차별적인가? 이러한 오류가 비뚤림을 유발하였는가? 효과에 어떤 방향으로 비뚤리게 하는가? 등이다.

3. 연구(분석)참가자 선정에서 발생하는 바이어스 위험

연구 대상자 선정과 관련된 바이어스는 대상자의 선택이 노출과 결과에 영향을 미치는 경우, 또는 추적 시간이 분석에서 제외되는 경우에 발생하게 된다. 예를 들어, 임산부의 엽산 섭취와 신생아 신경관 결손에 관한 연구에서는 유산아나 사산아의 경우 신경관 결손을 평가할 수 없으며, 이들은 신생아로 태어나지 못하게 된다. 따라서 신생아를 연구 대상으로 선택하는 것은 엽산 섭취(생존아가 엽산 섭취 가능성이 더 높음)와 신경관 결손(생존아가 신경관 결손 가능성이 더 낮음) 모두에 영향을 미치게 된다.

추적 시간의 일부가 분석에서 제외되거나 사건이 발생하지 않는 추적 기간(immortal time bias)을 갖는 경우, 이러한 바이어스의 위험은 일반적으로 유병환자에서 발생하게 된다. 예를 들어, 지역사회 조기 복귀 프로그램과 장기 입원의 효과를 비교하는 연구에서, 지역사회 조기 복귀 프로그램에 참여한 환자가 입원 기간 동안 사망한 경우, 이들은 지역사회 조기 복귀를 할 수 없으므로 이러한 환자가 제외되어 긍정적인 효과가 나타날 수 있다.

참가자 선정과 관련된 내용을 평가하기 위해서는 대부분의 참가자가 추적 관찰이 노출의 시작시점에서 시작되었는지? 분석된 추적관찰기간 동안 노출의 효과가 일정하다고 가정할 수 있는지? 참여자의 선택 기준이 노출시작 이후에 관찰된 특성으로 기인된 경우인지? 참가자 선정이 결과 또는 결과의 원인에 의해 영향을 받았을 가능성이 있는지? 등을 고려할 수 있다.

4. 노출 이후 개입으로 인한 바이어스 위험

해당 평가항목은 BOBINS-E에만 해당되는 평가항목이다. 노출이후 특정 중재가 노출기간 중 혹은 추적기간 중에 발생하며 특히 이전 노출과 결과에 영향을 미치는 경우를 의미한다. 예를 들어 석면에 노출된 사람들과 폐암발생 및 사망과의 관련성을 확인한 연구에서 석면에 노출된 사람들이 CT 검사를 받을 확률이 높았고, 이 검사가 폐암 사망 위험에 영향을 미친 경우가 있다. 이는 노출된 사람들이 노출되지 않은 사람들과 달리 특정 검사나 중재를 더 많이 받게 되는 상황에서 발생할 수 있다.

따라서 추적기간 중에 이전 노출로 기인하여 특정 개입을 받았는지? 만약 그렇다면 이 개입의 효과를 분석으로 교정했는지 등을 확인하게 된다.

5. 결측된 데이터로 기인된 바이어스 위험

많은 관찰연구에서 최초 연구대상에서 분석대상으로 참가자 선정 과정을 갖음에도 그 과정에 대한 기술을 생략하거나 혹은 노출여부, 관련 공변량의 여부, 결과확인 여부에 따라 대상자를 제외하기도 한다. 이러한 경우 제외된 대상자를 결측되었다고 평가하여야 한다. 또한 이러한 제외가 실제 결과값과 관련이 있는지에 대한 평가가 필요하다. 특히 양군의 결측율이 다르거나 혹은 결측사유가 다르거나 혹은 결측률이 결과발생에 비해 높은 비율을 갖는 경우 바이어스 위험이 발생할 가능성이 있다.

결측된 데이터는 노출, 교란변수, 결과변수 등 모든 관련 변수의 결측으로 인해 분석에서 제외되는 경우 해당 결측이 결과와 무관하게 발생되었다는 것에 대한 민감도 분석이 필요하며, 만약 무작위적으로 발생한 결측이라면 적절한 대치방법이 사용되었는지 확인해야 한다. 적절한 대치방법이 사용되었다고 하더라도 결측 자체가 무작위적이지 않은 경우는 대치방법 만으로는 연구결과의 바이러스를 보정하기는 어렵다. 따라서 결측사유에 대한 깊이 있는 평가가 필요하다.

6. 결과의 측정과 관련된 바이어스 위험

결과의 측정은 연구참가자 혹은 결과를 평가하는 평가자가 대상자의 노출에 대한 정보를 알았는지를 평가하는 것으로 노출그룹 또는 노출수준에 따라 확인방법이 달랐는지, 결과 평가자가 연구 참가자들의 노출력을 알고 있었는지 등에 대한 평가항목으로 구성되어 있다. 결과평가에 대한 바이어스는 상대적으로 코호트연구에서 더 영향이 크며, 일반적으로 무작위연구와 같이 계획에 따른 결과평가를 하지 않는 경우 보고자 하는 노출과 건강 결과와의 관련성이 있을 때 더 많은 검사 혹은 측정을 하게 되어 대

조군과는 다른 감시체계를 갖게 된다. 따라서 이 부분에 대한 주의가 필요하다.

또한 관심결과 판단과정의 객관성 여부에 따라 연구결과가 바이어스 될 가능성의 차이가 존재하게 된다. 관심결과에 대한 측정은 신뢰성과 타당성이 입증된 도구로서 측정되어야 하며 특정 장비나 검사 등으로 결과를 수집한 경우에도 측정장비의 정확도 등에 대한 평가가 필요하다.

7. 연구결과의 보고와 관련된 바이어스 위험

연구 결과의 선택이 효과의 크기나 방향과 관련이 있을 때 발생할 수 있으며, 이는 긍정적이지 않은 연구 결과를 누락시키거나 통계적으로 유의한 결과만을 보고하는 경우에 해당한다. 여러 측정값 중에서 추정효과의 크기에 따라, 분석 방법 측정 시점, 하위군 등의 선택이 이루어져 보고되었는지를 판단하는 것이다.

4 요약

역학적 논문의 비평적 평가를 위해서는 각 연구설계 혹은 주제에 적합한 평가도구를 선정하게 되고 연구주제와 관련된 평가기준을 선정하는 것이 필요하다.

예를 들어 중요한 기저 대상자의 특성 및 위험/예후요인 설정과 적절한 기준, 일차결과의 타당성, 결과측정과 필요 시 노출 측정의 맹검적용, 적절한 추적관찰기간 및 수용가능한 탈락률, 관심 노출, 결과, 교란요인 등의 측정방법의 타당성과 신뢰성 등이 해당된다.

관찰 연구의 특성을 주로 갖는 역학 연구는 무작위화되지 않았기 때문에, 코호트연구의 경우 노출 여부에 따른 군 간 비교 가능성이 유지되었는지를 평가하는 것이 필요하다. 또한, 환자-대조군연구에서는 관심 결과를 가진 대상과 그렇지 않은 대상을 제외하고, 다른 중요한 요인에서의 비교 가능성을 평가해야 한다. 또한 연구설계 특이성을 가진 바이어스로 환자-대조군연구에서 회상바이어스, 설문조사를 시행하는 경우 면접자 바이어스, 코호트연구에서 결과확인 바이어스 등 추가 고려사항이 있는지도 파악해야 한다. 관찰연구는 대부분 모든 경우 교란요인이 발생할 수 있으며, 이러한 교란요인을 통제하기 위해 적절한 통제 방법이 연구설계 혹은 분석과정에서 사용되었는지에 대한 평가가 필요하다. 연구에서 발생 가능한 바이어스를 확인하기 위해서는 연구를 평가하기 이전 가장 이상적 조건에서의 연구설계를 미리 계획하고 실제로 연구를 수행하기위한 잠재적인 장애요인을 도출해 보는 것이 필요하며,

이러한 방법은 해당 연구주제의 문제점과 극복방안을 이해하는데 도움이 될 수 있다.

연구를 비판적으로 읽고 해석하는 것은 잘 수행된 연구와 그렇지 않은 연구를 구분하기위해서라기보다는 바이어스가 연구결과에 미치는 영향의 정도와 방향을 추정함으로써 해당 연구결과를 보다 잘 이해하기 위해서이다.

참고 문헌

1. AHRQ Methods for Effective Health Care. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008.
2. Grimes DA, Schulz KF. Bias and causal associations in observational research. Lancet. 2002;359(9302):248-52.
3. Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. Lancet. 2002;359(9303):341-5.
4. Higgins JPT, Morgan RL, Rooney AA, et al. A tool to assess risk of bias in non-randomized follow-up studies of exposure effects (ROBINS-E). Environ Int. 2024;186:108602.
5. Parr CL, Hjartåker A, Laake P, et al. Recall bias in melanoma risk factors and measurement error effects: a nested case-control study within the Norwegian Women and Cancer Study. Am J Epidemiol. 2009;169(3):257-66.
6. Schulz KF, Grimes DA. Case-control studies: research in reverse. Lancet. 2002;359(9304):431-4.
7. Seo HJ, Kim SY, Lee YJ, et al. RoBANS 2: A Revised Risk of Bias Assessment Tool for Nonrandomized Studies of Interventions. Korean J Fam Med. 2023;44(5):249-60.
8. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.
9. Wang X, Cheng Z. Cross-Sectional Studies: Strengths, Weaknesses, and Recommendations. Chest. 2020;158(1s):S65-S71.

제2절 체계적문헌고찰 논문의 해석

1 체계적문헌고찰 정의와 특성

체계적문헌고찰(systematic review)은 특정 연구 질문에 대한 기존 문헌을 체계적이고 명확한 방법론을 통해 종합·분석하는 연구 방법이다. 다양한 학문 분야에서 사용되며, 연구 목적과 방법에 따라 정의가 조금씩 다를 수 있다. 체계적문헌고찰의 연구방법과 관련된 가장 대표적인 국제 단체인 코크란 연합(Cochrane Collaboration)은 "명확한 연구 질문을 설정한 후, 사전 정의된 방법론을 사용하여 관련 문헌을 체계적으로 검색하고 선정 및 배제 기준을 적용하며, 질 평가 및 데이터 분석을 수행하여 근거 기반 결론을 도출하는 연구 방법."이라고 정의하고 있으며, 체계적문헌고찰 논문의 보고지침인 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)에서는 체계적 문헌고찰을 "연구 질문을 중심으로 관련 문헌을 체계적으로 검색, 선택, 평가 및 종합하는 과정으로, 연구의 타당성과 재현성을 보장하기 위해 명확한 프로토콜을 따르는 연구 방법"이라고 정의하고 있다.

이렇듯 체계적문헌고찰은 명확한 연구 질문을 기반으로, 관련된 모든 연구를 체계적이고 엄격한 방법으로 검색하고 평가하여 분석하는 과정이라고 요약할 수 있다. 이 과정은 구조화되어 있으며, 여러 연구들의 결과를 종합하여 결론을 도출하는 데 중점을 둔다. 체계적문헌고찰의 주요 특징은 다음과 같다

① 명확한 연구 질문: 대답가능한 구체화된 연구질문으로 일반적으로 PICO의 형식으로 구조화한다.
② 포괄적 문헌검색: 가능한 한 모든 관련 연구를 체계적으로 검색하고 포함하도록 한다.
③ 엄격한 선정 기준: 고찰에 포함될 논문과 제외될 논문을 명확히 구분하는 포함/배제 기준을 사전에 설정하여 두 명 이상의 연구자의 독립적 선정원칙을 따른다.
④ 바이어스 위험 평가: 포함된 각 연구의 바이어스 위험을 평가한다. 단 이때 각 연구의 설계를 기반으로 적절한 도구를 선택하여 적용한다.
⑤ 데이터 통합: 연구 결과를 메타 분석(Meta-analysis) 등을 통해 통합하여 결론을 도출한다. 연구 결과를 합산할 때 합산할 수 있을 정도의 충분한 동질성에 대한 가정이 필요하다. 메타분석은 기

본적으로 2개 군을 비교하는 방식으로 제시되나, 2개 이상의 치료를 비교 분석하기 위한 네트워크메타분석과 개별 환자의 원자료를 포함하는 메타분석방법도 사용 가능하다.

⑥ 투명성: 연구 방법, 검색 과정, 평가 기준 등이 명확히 제시되어 재현 가능한 방식으로 진행되며 이를 제시하여야 한다.

2 체계적문헌고찰의 결과 해석 시 방법론적 고려사항

체계적문헌고찰의 평가도구 중 일반적으로 사용되고 있는 도구는 AMSTAR 2(A Measurement Tool to Assess Systematic Reviews 2)로 2007년 개발된 도구의 업데이트 버전이다. 기존 도구가 무작위대조임상시험의 체계적문헌고찰을 평가하는 목적으로 개발된 반면, 중재에 대한 비무작위 연구의 체계적문헌고찰을 추가하여 보다 광범위하게 적용되는 도구로 개발되었다는 점에 차이점이 있다. 총 16개의 평가항목으로 구성되어 있으며 이 중 아래의 7개의 평가항목은 체계적문헌고찰의 결과에 영향을 많이 미칠 수 있는 중요 항목으로 구분하여 제시하고 있다. 이들 7개의 평가항목을 기반으로 체계적문헌고찰 문헌을 해석하는 데 고려할 점으로 요약하였다.

표 9-2-1. AMSTAR 평가항목 중 중요하게 고려되어야 할 항목 7개

체계적고찰 시작 전에 프로토콜 등록(항목 2)
문헌검색의 적절성(항목 4)
개별 연구를 제외하는 정당성(항목 7)
고찰에 포함된 개별 연구에 대한 바이어스 위험 평가(항목 9)
메타분석 방법의 적절성(항목 11)
고찰결과 해석 시 바이어스 위험에 대한 고려(항목 13)
출판 바이어스의 존재 및 가능성에대한 영향 평가(항목 15)

1. 체계적문헌고찰의 시작 전에 프로토콜 등록 및 개별연구를 제외하는 정당성

보고바이어스는 연구결과의 특성에 따라 연구결과가 선택적으로 보고 혹은 배포되는 것을 의미한다. 보고바이어스는 체계적문헌고찰의 대상이 되는 일차연구에서도 발생가능하며, 체계적문헌고찰 과정에서도 발생 가능한 문제로 연구자의 의도에 따라 결과지표의 탈락, 변경이 일어나거나, 혹은 개별 연

구의 선택의 배제가 일어나기도 한다. 이러한 보고바이어스를 예방하기 위한 방법 중 하나가 연구시작 전 프로토콜을 사전에 작성하고 이에 따라 연구를 수행하는 것이다. 체계적문헌고찰도 다른 유형의 연구와 마찬가지로 모집단, 중재 또는 노출, 관심결과를 포함한 연구질문을 생성하여 프로토콜을 작성해야 하며 대표적인 체계적문헌고찰의 프로토콜 온라인 등록 사이트는 PROSPERO (International prospective register of systematic reviews)이다. 코크란 리뷰가 프로토콜을 먼저 작성하고 이후 리뷰를 작성하는 단계로 이루어져 있는 것도 이러한 보고바이어스의 예방 및 연구자의 의도로 인한 연구결과 생성의 문제를 최소화하기위한 방안 중 하나로 평가될 수 있다.

2. 출판바이어스의 존재 및 가능성에 대한 영향 평가 및 문헌검색의 적절성

체계적문헌고찰의 가장 중요한 연구목표는 연구질문과 관련된 모든 데이터를 식별하는 것이다. 데이터의 검색은 저널 데이터베이스(예: PUBMED), 임상시험등록부(예: Clinicaltrial.gov) 등을 통해 이루어지며, PubMed 및 EMBASE 와 같은 전자 데이터베이스가 게시된 기사를 식별하는 것을 훨씬 더 쉽게 만들었지만 체계적 고찰은 여전히 누락된 연구의 미포함의 가능성이 있으며 이러한 바이어스의 한가지 원인은 출판바이어스이다. 여러 연구에 따르면 수행된 연구의 절반만이 출판되며, 통계적으로 유의미한 결과는 출판될 가능성이 더 높고, 유의미하지 않은 결과가 있는 연구보다 평균적으로 1년 일찍 출판된다고 보고하고 있다. 통계적으로 유의성에 따라 연구의 보고가 누락된다면 체계적 고찰의 결과는 보고된 긍정적이고 유의한 결과로 과장될 가능성을 가지게 된다.

비공개로 인한 근거 선택의 제한은 파악하기 어려울 수 있지만, 이러한 문제를 해결하기 위해 가능한 모든 기존 데이터를 검색하는 데 사용할 수 있는 전략이 여러 가지 제시되고 있다. 임상시험 등록부, 규제 기관 웹사이트, 학술 대회 초록을 검색하여 미공개 연구나 출판물에서 선택적으로 생략되었을 수 있는 결과를 파악할 수 있다. 세계보건기구(WHO)에서는 임상시험 검색 포털(http://apps.who.int/trialsearch)을 통해 여러 임상시험 등록부를 검색할 수 있게 도와주고 있다. 또한 회색문헌을 보다 광범위하게 검색하기 위한 다양한 웹사이트들이 서비스를 제공하고 있다.

출판바이어스를 평가하는 방법으로 그래픽 및 통계적 방법이 개발되어 있으나 일반적으로 연구수가 작을수록 검정의 한계를 가지고 있다. 특히 10개 미만의 연구가 대상인 경우 사용하지 않는 것을 권고하고 있으므로 해당 방법을 통한 출판바이어스에 대한 엄격한 평가는 한계가 있다. 따라서 연구주제에 따른 출판바이어스의 가능성이 존재할 수 있는지에 대한 임상적 판단이 필요하며, 회색문헌의 포함을 확대시키는 등 엄격한 문헌검색이 중요하다.

3. 포함된 개별연구에 대한 바이어스 위험 평가 및 결과해석 시 바이어스 위험에 대한 고려

체계적문헌고찰은 연구주제에 관련된 모든 연구의 데이터를 모아 합산하는 특성을 가진 연구설계이다. 그러므로 체계적문헌고찰의 결과는 모아진 연구데이터를 기반으로 하며, 모아진 개별연구가 바이어스 위험이 높다면 체계적문헌고찰의 결과 역시 바이어스 위험이 높을 수밖에 없는 근본적인 한계를 갖는다. 따라서 체계적 고찰에 포함된 각 개별 연구의 주요 바이어스 위험에 대해 평가하고 이러한 평가결과가 연구결과를 해석할 때 반영되어야 한다. 개별연구의 바이어스 위험은 연구의 설계, 중재의 제공, 결과의 평가, 연구결과의 선택과 보고 등 연구 전반에 걸쳐 발생할 수 있다. 예를 들어 무작위 시험의 경우 연구자 혹은 참여자의 맹검이 이루어지지 않는 경우 연구자 혹은 참여자의 의도가 반영되어 중도탈락 혹은 중재 이행의 불균형, 혹은 결과지표의 주관적 판단 개입 등이 일어날 수 있다. 맹검이라는 단일의 요소가 연구 전반에 영향을 미치게 되는 예이다. 따라서 연구설계에 적합하게 개발된 도구를 선정하여 개별 연구를 평가하는 것이 필요하며, 평가도구 중 정량적 척도를 사용하는 것은 항목별 가중치에 대한 불명확함이 존재하므로 각 항목에 대한 정성적 평가를 사용하는 것을 코크란 연합에서는 추천하고 있다.

4. 메타분석의 적절성

메타분석은 두 개 이상의 독립된 연구결과를 통계적으로 결합하는 것으로 다양한 연구의 효과추정치를 가중 평균하여 합산하는 방식을 따른다. 여러 연구결과를 합산하는 것이므로 연구결과를 합산하는 과정에서 연구간 변동(이질성), 개별 연구의 바이어스 위험, 체계적 고찰 과정에서의 보고바이어스 등에 대한 신중한 고려가 필요하나, 연구결과의 정밀도를 향상시키고 상충되는 연구결과의 논란을 해결할 수 있는 장점을 가지고 있다.

기본적으로 메타분석의 적절성을 평가하기 위해서는 메타분석의 기본 원칙을 준수하는 것이 필요하다. 기본적으로 AMSTAR 2에서 제시하고 있는 지표는 연구결과를 합산할 만큼 충분한 동질성, 적절한 가중치 적용방식의 선택, 연구설계가 다른 경우 구분하여야 하며 비무작위 연구의 경우 교란요인을 통제한 보정된 연구결과를 사용해야 함을 제시하고 있다. 또한 바이어스위험이 높은 연구가 포함되는 경우, 누락된 데이터가 존재하는 경우 이에 대한 적절한 처리 및 결과해석에서의 반영 여부도 메타분석의 적절성을 평가하기 위해 고려해야 할 부분이다.

3 체계적문헌고찰의 분석결과 해석

체계적문헌고찰은 다른 연구설계에 비해 정형화되어 있는 연구결과 제시방법이 있다. 이에 대한 해석을 위해 2022년 발표된 코크란 리뷰인 〈Prophylactic anticoagulants for non-hospitalised people with COVID-19〉의 결과를 인용하였다.

1. 검색식

검색은 체계적문헌고찰에서 가장 중요한 연구방법 중 하나이다. 일반적으로 체계적문헌고찰 논문 연구방법 부분에 첫 번째로 기술된다. 코크란리뷰의 경우는 다른 저널에 비해 보다 자세히 기술되어 있다. 검색 시 확인해야 할 사항으로는 최소 두 개 이상의 전자데이터베이스 활용, 꼭 필요한 경우가 아니라면 제한 검색 해제, 출판되지 않은 연구를 포함하기 위한 노력, 관련 전문가 자문, 수기 검색, 최신 연구 반영(최근 24개월 이내) 등이 있다. 아래의 예시(그림 9-2-1)를 살펴보면 언어, 출판연도, 출판상태에 따른 제한을 하지 않음을 보고하고 있고 다양한 데이터베이스를 포함하여 검색하였다. 또한 계획서 등록원, 참고문헌목록 검토, 저자와의 연결 등 여러 방식의 수기검색을 수행하였음을 보고하였다.

The Cochrane Vascular Information Specialist conducted systematic searches of the following databases for RCTs and controlled clinical trials without language, publication year, or publication status restrictions:

- Cochrane Vascular Specialised Register via the Cochrane Register of Studies (CRS-Web);
- Cochrane Central Register of Controlled Trials (CENTRAL; 2022, Issue 4) via the Cochrane Register of Studies Online (CRSO);
- Cochrane COVID-19 Study Register via the CRSO;
- MEDLINE (Ovid MEDLINE Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE Daily and Ovid MEDLINE); Embase Ovid;
- CINAHL (Cumulative Index to Nursing and Allied Health Literature); and
- LILACS (Latin American and Caribbean Health Science Information database) (via Virtual Health Library).

controlled clinical trials (as described in Chapter 4 of the Cochrane Handbook for Systematic Reviews of interventions) (Lefebvre 2022). Search strategies for the major databases are provided in Appendix 1.

We searched the following trial registries:

- World Health Organization International Clinical Trials Registry Platform (www.who.int/clinical-trials-registry-platform); and Clinical Trials.gov (clinicaltrials.gov).
- The most recent searches were carried out on 18 April 2022.

Searching other resources

We checked the reference lists of all included studies and any relevant systematic reviews identified for additional references to studies. We examined any relevant retraction statements and errata for the included studies. We contacted the authors of the included studies for any possible unpublished data. We contacted field specialists to enquire about relevant ongoing or unpublished studies.

그림 9-2-1. 검색데이터베이스 예시

검색식은 일반적으로 부록에 제시하게 된다. 검색식은 아래의 예시(그림 9-2-2)처럼 대상이 된 모든 데이터베이스별 검색일자, 자연어와 통제어휘를 모두 포함한 검색식이 제시되어야 한다.

Source	Search strategy
1. Medline (Ovid MEDLINE® Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE® Daily and Ovid MEDLINE® 1946 to present (Date of most recent search: 18 April 2022)	1. exp COVID-19/ 2. COVID-19.ti,ab 3. COVID19.tab 4. 2019 novel coronavirus infection.ti.ab. 5. coronavirus disease 19tiab 6. 2019 novel coronavirus disease. ti,ab 7. coronavirus disease 2019.ti,ab. 8. exp Coronavirus/ 9. Coronavirus*.ti,ab. 10. Deltacoronavirus*.ti,ab. 11. Delta-coronavirus*.ti,ab. 12. or/1-11 13. exp Anticoagulants/ 14. (anticoagul or anti-coagu*).ti,ab. 15. exp Heparin/ 16. heparin*. ti,ab/

Source	Search strategy
2. EMBASE via Ovid (Date of most recent search: 18 April 2022)	1. exp COVID-19/ 2. COVID-19.ti,ab. 3. COVID19.ti,ab. 4. 2019 novel coronavirus infection.ti.ab. 5. coronavirus disease 19.ti,ab. 6. 2019 novel coronavirus disease.ti,ab 7. coronavirus disease 2019.ti,ab. 8. exp Coronavirus/ 9. Coronavirus ti,ab 10. Deltacoronavirus.ti,ab. 11. Delta-coronavirus.ti,ab. 12. or/1-11 13. Anticoagulants/ 14. (anticoagul or anti-coagu").ti,ab. 15. exp Heparin/ 16. heparin ti,ab.

그림 9-2-2. 데이터베이스별 검색식 예시

2. 연구흐름도

체계적문헌고찰은 대부분의 경우 연구흐름도(PRISMA flow diagram)를 첫 번째 그림으로 제시하게 된다. 검색과 연구선택의 절차를 보여주는 그림이다. 검색과 연구선택의 절차가 적절한 지 여부를 판단하게 되며 특히 전체 원문을 보고 선택에서 배제된 연구의 목록과 이유가 중요하다. 예시(그림 9-2-3)에서는 18개의 연구(19개의 보고자료)를 제외하였다고 설명하고 있다, 코크란리뷰의 경우 배제된 논문의 목록과 이유를 부록에 제시하므로 확인이 필요하다.

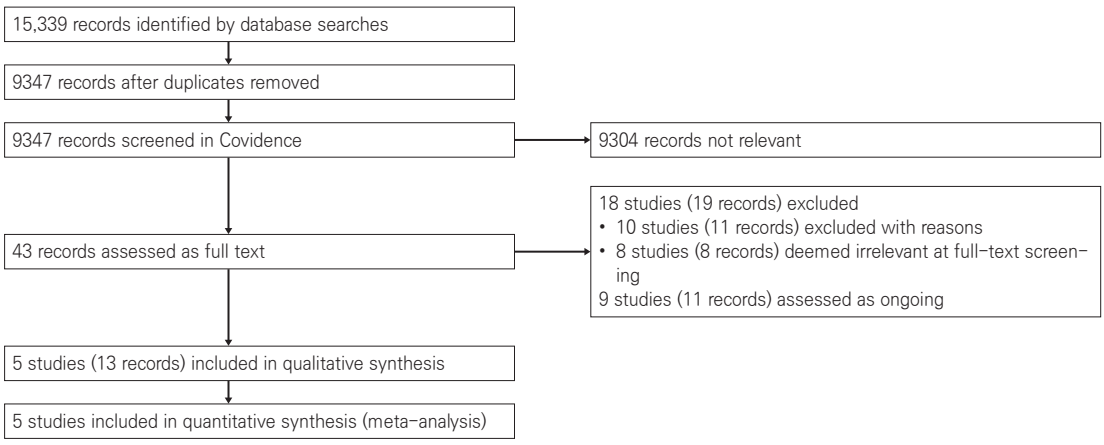


그림 9-2-3. 문헌선택의 흐름도 예시

본 예시에서 5개의 연구와 13개의 보고자료가 포함되었으며 이들 모두 정량적 분석에 포함되었음을 제시하고 있다. 연구(study)와 보고(record)의 차이점은 동일한 대상자로 수행된 연구가 다른 형태(포스터, 논문, 기사 등), 다른 추적기간, 다른 결과 등 다양한 이유로 여러번의 보고가 이루어진 경우를 구분하는 것으로 체계적문헌고찰은 동일한 대상자의 연구가 반복적으로 여러번 보고된 경우에 하나의 결과를 선택하여 사용하도록 하고 있다. 하나의 결과를 사용하는 경우 사전 원칙이 필요하며, 일반적으로는 최종보고된 연구결과를 사용하지만, 연구결과에 따라 추적기간 별로 혹은 결과가 다른 경우는 각각을 모두 사용할 수도 있다. 하지만 하나의 메타분석 내에 동일한 연구대상이 두 번 이상 포함되지 않도록 한다.

3. 바이어스위험 평가표

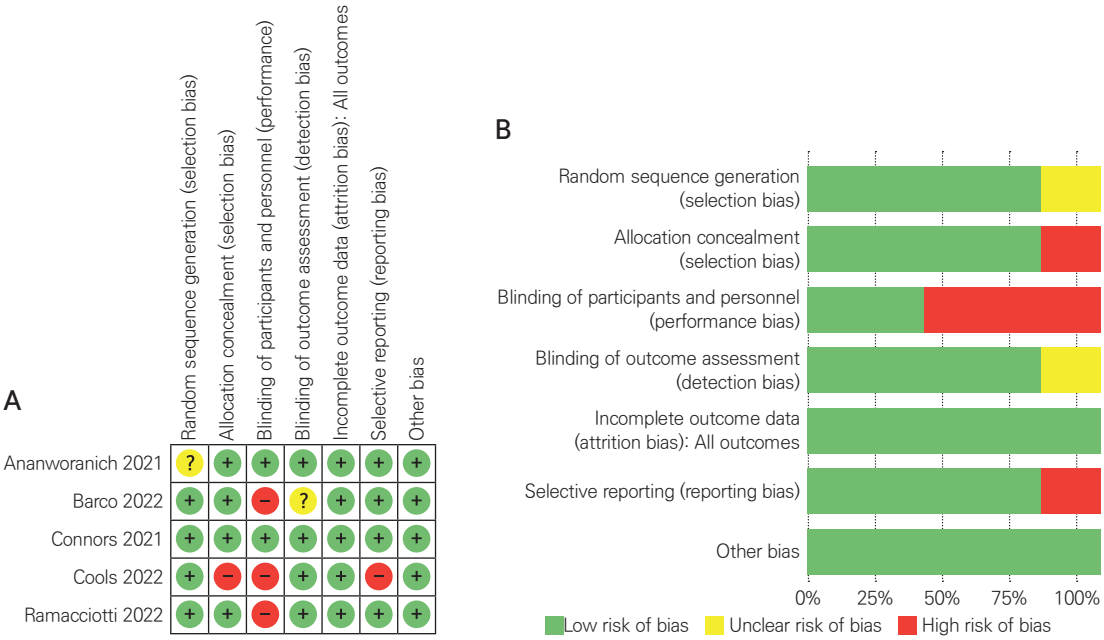


그림 9-2-4. 개별연구의 바이어스 평가결과 제시 예시

바이어스 위험 평가결과는 다음 그림과 같은 형태로 표현하여 제시하는 것이 일반적이다. 예시자료(그림 9-2-4)는 무작위연구를 대상으로 하고 있으며 평가도구는 코크란그룹에서 개발된 무작위대조실험 연구에 대한 바이어스위험 평가도구의 과거버전이다. 현재는 5개 항목으로 구성된 새로운 버전의 도구가 개발되어 사용되고 있다. 그림의 형태는 2가지로 모두 제시할 수 있으며 (A)의 경우는 개별 연구별

개별 평가항목별 평가결과를 제시한 형태이고 (B)의 경우는 각 평가항목별 합산결과를 제시한 표이다. 두 가지 형태 중 무엇을 선택하여도 상관없으나 개별연구별 평가결과를 제시하는 것이 보다 추천되는 형태이다.

다음의 결과는 다음과 같이 해석할 수 있다.

> 총 5개의 연구가 포함되었으며 참여자와 연구자의 맹검에 대한 항목에서 포함된 연구의 3개에서 바이어스의 위험이 높았다. 따라서 해당 항목이 다른 항목에 비해 바이어스될 위험이 높았다. Cools 2022 연구의 경우 배정은폐 및 연구자, 참여자 맹검, 연구결과 선택의 항목에서 바이어스될 위험이 높았으며, 전반적 바이어스 위험이 높다. 그 외 Barco 2002 연구는 결과평가자의 맹검에서의 불명확함이 있으며, Ananworanich 2021 무작위방법에서의 불명확함이 있었으나 그 외 항목의 경우 대부분의 연구는 바이어스될 가능성은 낮았다. 따라서 전체적으로 환자가 보고된 연구 혹은 연구자의 결정이 필요한 연구결과는 바이어스될 가능성이 있다.

4. 메타분석결과표

메타분석 결과는 통계프로그램마다 조금씩 차이가 있으나 기본적인 형태는 모두 동일하다. 다음 그림은 이분형자료에 대한 메타분석 결과인 숲그림이다. 레브만이라는 프로그램을 통해 분석된 결과이며 바이어스 위험을 같이 제시할 수 있다는 점에서 장점을 가지고 있다. 숲그림을 읽는 방법은 맨 좌측에 포함된 연구의 ID(일반적으로 첫 저자의 이름과 출판연도), 목표 중재의 사건과 대상자수, 대조군의 사건과 대상자수, 가중치, 개별연구의 효과 추정치가 제시된다. 그림에서 효과의 추정치로 제시되는 네모의 크기는 개별연구의 가중치를 의미하며 일반적으로 가중치는 대상자수, 사건수, 분산 등을 통해 결정된다.

메타분석의 이질성은 Q 통계량과 I2값을 통해 평가한다, I2은 50%를 기준으로 작은 경우 중등도 이하의 이질성을 의미하나 사건수, 연구의 대상자수의 수, 분산 등에 영향을 많이 받으므로 신뢰구간의 겹침정도, 점추정치의 유사성을 함께 확인하는 게 필요하다.

아래의 예시는 입원에 대한 위험도이다. 해석은 다음과 같다.

> 위약군 혹은 무치료군에 비해 항응고제를 투여받은 대상자의 입원에 대한 상대위험도는 RR=0.94(0.51-2.31)로 감소되었으나 통계적으로 유의하지 않고 연구의 이질성은 없다(I^2=0%)

개별연구 추정치에 대한 가중치와 바이어스 위험평가 결과를 비교해 보는 것이 중요하며 상대적으로 바이어스 위험이 높은 연구의 결과가 전체 결과에 기여한 바를 파악한다. 다음 예시(그림 9-2-5)에서 전체 연구결과의 67.2%를 Barco 2022의 연구결과가 기여하였으며, 해당 연구는 포함된 다른 두 개 연구에 비해 바이어스위험이 상대적으로 높다. Cools 2022의 결과는 바이어스 위험이 높아서 연구결과에 포함되지 않음을 제시하고 있다.

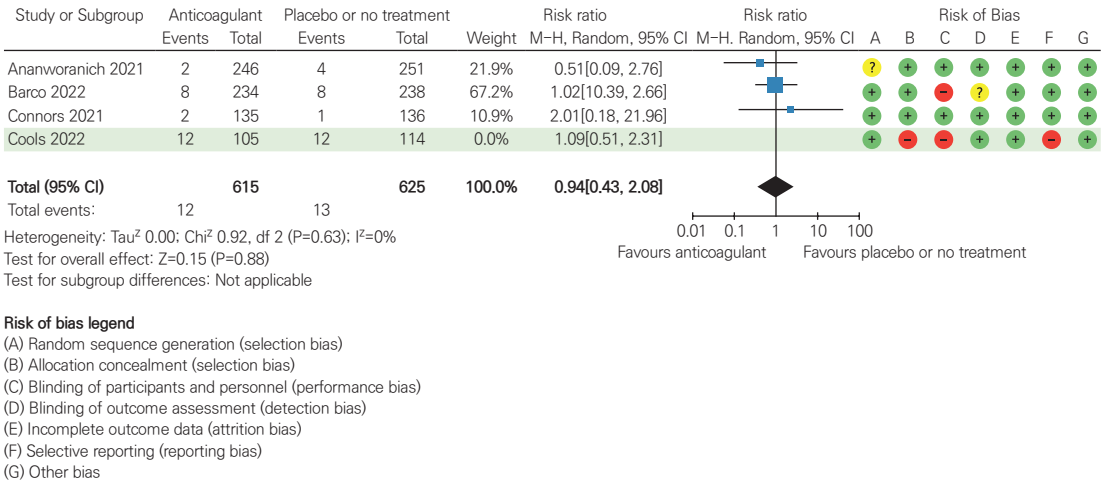


그림 9-2-5. 메타분석 숲그림 예시

5. 출판편향에 대한 판단표

출판바이어스를 평가하는 분석방법으로 보고된 연구결과의 효과와 정밀성의 관련성을 그림으로 평가하는 방법인 깔대기그림(funnel plot)이다. 연구의 정밀도(대상자수)가 증가할수록 연구결과는 합산된 추정치의 중심에 근접하게 되고 정밀도가 낮은 연구의 경우는 연구결과는 표본추출과 관련된 문제로 추정치에서 떨어지게 되는 특성을 갖는다. 이를 표현한 것이 깔대기그림이다. 연구의 추정치를 중심으로 양군이 대칭성이 있는 경우 출판바이어스의 위험이 낮다고 평가하게 되고 만약 아래의 예시(그림 9-2-6)처럼 일정 부분의 결과가 미보고되었다면 출판바이어스가 의심된다고 평가할 수 있다. 단 출판바이어스의 경우는 10개 이상의 연구에서 평가가능하므로 그 이하의 연구에서의 검정력은 없으며, 10개 이상의 경우라고 하더라도 검정력이 낮아 출판바이어스에 대한 통계검정 결과가 유의하지 않는 것이 출판바이어스가 없다고 결론짓기는 어렵다.

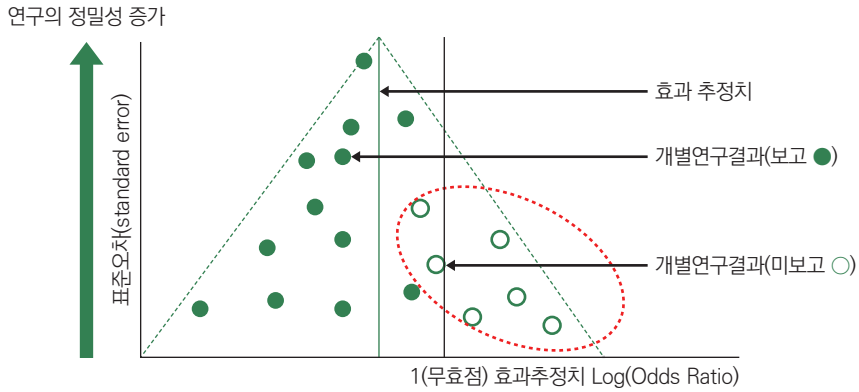


그림 9-2-6. 출판바이어스 평가 그림

6. 결과요약표

체계적문헌고찰의 결과는 다음의 결과요약표(summary of findings, SoF; 표 9-2-2)로 제시될 수 있다. 결과요약표는 각 연구결과에 대한 요약된 정보를 제시하는 것으로 기저위험도(위약군, 무처치군의 위험), 대응위험도(항응고제의 위험), 상대위험도(RR), 연구대상자 및 연구 수, 근거의 확신도 평가결과가 한번에 요약되어 있는 표이다.

표 9-2-2. 연구결과요약표 예시

Anticoagulant versus placebo or no treatment for non-hospitalised people with COVID-19

Patient or population: non-hospitalised people with COVID-19
Setting: outpaitent
Intervension: anticoagulant
Comparison: placebo or no treatment

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	No of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with placebo or no treatment	Risk with anticoagulant				
All-cause mortality Follow-up: from 30 to 90 days	Study population		RR 0.36 (0.04 to 3.61)	1,777 (5 RCTs)	⨁⨁◯◯ Lowa,b	
	6 per 1,000	2 per 1,000 (0 to 20)				

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	No of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with placebo or no treatment	Risk with anticoagulant				
Venous thromboembolism Follow-up: from 30 to 90 days	Study population		RR 0.36 (0.16 to 0.85)	1,259 (4 RCTs)	⊕⊕⊕◯ Moderatea, c	NNTB =50
	31 per 1,000	11 per 1,000 (5 to 26)				
Major bleeding Follow-up: from 30 to 90 days	Study population		RR 0.36 (0.01 to 8.78)	1.77 (5 RCTs)	⊕⊕◯◯ Lowa,b	
	1 per 1,000	0 per 1,000 (0 to 10)				
Deep vein thrombosis Follow-up: f rom 30 to 90 days	Study population		RR 1.02 (0.30 to 3.46)	1,009 (3 RCTs)	⊕⊕⊕◯ Moderatea, c	
	10 per 1,000	10 per 1,000 (3 to 34)				

* The risk in the intervention group (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI)
CI: confidence interval; NNTB: number needed to treat for an additional beneficial outcome; RCT: randomised controlled trial; RR: relative ratio

GRADE Working Group grades of evidence

High certainty: we are very confidence that the effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate; the true effect is likely to be close to the estimate of the effect, but there is possibility that it is substantially different.

Low certainty: our confidence in the effect estimate is limited; the true effect may be substantially different from the estimate of the effect.

Very low certainty: we have very little confidence in the effect estimate; the true effect is likely to be substantially different form the estimate of effect.

위의 결과 중 첫 번째 결과인 모든 원인의 사망에 대한 결과는 다음과 같이 해석될 수 있다.

효과 추정치

치료를 하지 않거나 위약군은 1000명 중 6명이 사망하나(기저위험도) 항응고제를 투여한 군에서는 1,000명 중 2명이 사망한다(대응위험도), 상대위험은 0.36(95% CI:0.04-3.61)이며 5개의 무작위 대조군연구이며 포함된 대상자수는 1,777명이고 결과는 30일에서 90일까지의 기간 중에 발생한 사건에 대한 추정요약이다.

근거의 확신도

Low a, b: 근거의 확신도는 낮음이며 각주로 낮음인 이유에 대해 설명하고 있다. 연구에 포함된 개별연구의 바이어스 위험은 선택, 수행, 보고바이어스에서 높은 위험을 가지나 민감도 분석결과 연구결과에 영향을 미치지 않았으므로 하향조정하지 않고 다만 사건수가 작고, 신뢰구간이 넓어서 비정밀성 항목에서 두 단계 하향조정하여 낮음으로 평가하였다.

7. 근거의 확신도

근거의 확신도는 이전에는 근거수준(level of evidence)으로 사용되던 용어이나 현재는 근거의 확신도(certainty of evidence)로 사용되고 있다. 근거의 확신도는 GRADE 그룹이 제시한 메타분석의 근거의 신뢰도에 대한 평가방법이다. 코크란리뷰도 2011년부터는 GRADE 그룹이 제시한 근거의 확신도를 고찰결과에 제시하고 있다.

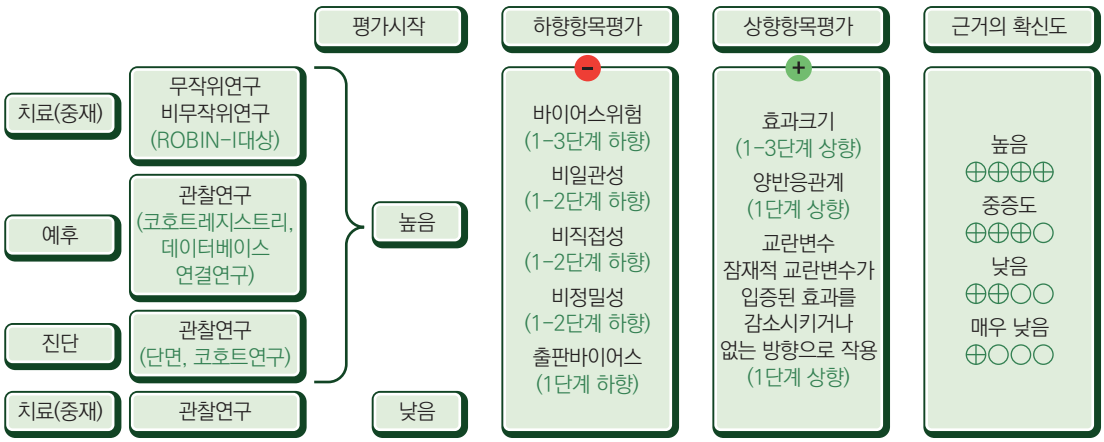


그림 9-2-7. 근거의 확신도 연구별 평가기준 및 과정

근거의 확신도는 메타분석을 하는 과정에서 생길 수 있는 여러 고려사항을 명료하게 정리한 내용으로 근거의 확신도를 상향시킬 수 있는 3가지 항목과 하향시킬 수 있는 5가지 항목을 제시하고 있다. 평가결과는 4가지(높음, 중등도, 낮음, 매우낮음)로 구분하며 기본적으로 연구설계를 기반으로 무작위연구의 경우 높음에서 시작, 비무작위 연구의 경우 낮음에서 시작하여 항목별 평가결과에 따라 한두단계 하향 혹은 상향조정을 하게 된다. 하지만 연구의 주제가 치료가 아닌 진단 혹은 예후의 경우는 연구설계와 상관없이 진단연구는 진단적 불확실성을 갖는 대상에서 수행했을 때 높음에서 시작하게 되며, 예

후연구 중 대표성이 있는 코호트연구인 경우 높음으로 시작하게 되는 것처럼 연구의 주제에 따른 적용 방법이 일부 차이가 존재한다.

하향조정을 하는 5가지 중요한 고려사항은 개별연구의 바이어스 위험 평가결과, 비직접성, 비동질성, 비정밀성, 출판바이어스이며, 상향조정하는 3가지 항목은 교란요인으로 인한 효과의 크기 감소된 경우, 양반응 효과, 큰 효과이다. 일반적으로 하향조정을 한 경우 상향조정을 하지 않는 것이 원칙이므로 상향조정은 비무작위연구의 경우가 대부분 대상이 된다. GRADE그룹에서 제시하는 근거의 확신도를 평가하는 과정을 요약하면 다음(그림 9-2-7)과 같다.

4 요약

체계적문헌고찰은 관련된 주제의 모든 연구를 찾아 평가하고 연구결과를 합산하는 연구방법이다. 연구결과의 검정력을 높이는 연구방법이으로 여러 연구의 결과를 하나의 결과로 합산하여 연구결과를 효율적이며 직관적으로 이해할 수 있게 하는 장점을 가지고 있다. 그러나 도출된 결과를 올바르게 해석하기 위해서는 연구의 주제와 대상에 따라 적절한 연구설계, 관심효과 추정치에 대한 분석 방법과 결과를 이해하는 것이 필요하다. 또한 포함된 연구의 바이어스 위험을 평가하는 방법은 연구설계별 적절한 도구를 선택하는 것뿐만 아니라 연구주제 특성에 따른 바이어스 위험을 평가하는 것이 필요하다. 따라서, 연구주제에 대한 충분한 지식을 가지고 있지 않은 경우 연구결과를 이해하거나 해석하기 어려울 수 있으며, 잘못된 연구결과나 결론이 도출될 가능성을 배제할 수 없다. 기본적인 일차연구에 대한 이해가 선행될 필요가 있다.

해당 관련 모든 연구를 모았다는 것 자체가 해당 연구의 신뢰성이 증가시키는 것은 아니며, 포함된 개별연구 혹은 체계적문헌고찰 과정에서의 바이어스가 있다면 연구결과는 여전히 신뢰성이 낮을 수밖에 없다. 또한 개별연구 결과를 요약하는 과정에서 많은 정보의 누락이 발생함으로 오히려 결과에 대한 비판적 평가가 제한될 수도 있다. 따라서 체계적문헌고찰을 읽고 해석할 경우 포함된 개별연구를 통한 추가적인 정보를 확인할 필요성이 있으며, 포함된 연구의 가중치 합산이므로 동일한 체계적고찰 내에서도 결과마다 근거의 확신도는 다를수 있다는 점을 기억할 필요가 있다.

참고 문헌

1. Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64(4):401–6.
2. Brozek JL, Akl EA, Jaeschke R, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. Allergy. 2009;64(8):1109–16.
3. Cochrane Training [Internet]. Cochrane Handbook for Systematic Reviews of Interventions Version 6.5, 2024 [Cited 2024 Dec 30]. Available from: https://training.cochrane.org/handbook/current.
4. Dwan K, Gamble C, Williamson PR, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias – an updated review. PLoS One. 2013;8(7):e66844.
5. Foroutan F, Guyatt G, Zuk V, et al. GRADE Guidelines 28: Use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks. J Clin Epidemiol. 2020;121:62–70.
6. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction–GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64(4):383–94.
7. Hopewell S, Clarke M, Lefebvre C, et al. Handsearching versus electronic searching to identify reports of randomized trials. Cochrane Database Syst Rev. 2007;2007(2):Mr000001.
8. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339:b2700.
9. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535.
10. National Institute for Health and Care Research [Internet]. PROSPERO(International prospective register of systematic reviews) [Cited 2024 Dec 30]. Available from: https://www.crd.york.ac.uk/prospero/.
11. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
12. Santos BC, Flumignan RL, Civile VT, et al. Prophylactic anticoagulants for non-hospitalised people with COVID-19. Cochrane Database Syst Rev. 2023;8(8):Cd015102.
13. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015;350:g7647.
14. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:j40.
15. The GRADE working group [Internet]. GRADE Handbook. [cited 2024 Dec 30]. Available from: https://gdt.gradepro.org/app/handbook/handbook.html.
16. Zorzela L, Loke YK, Ioannidis JP, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. BMJ. 2016;352:i157.

제3절 감시분석보고서의 작성

1 개요

1. 감염병 감시분석보고서 작성의 목적과 내용

역학은 인구집단에서 질병의 분포 양상을 파악하고, 결정요인을 구명하여 질병의 예방과 관리에 활용하는 학문으로 크게 기술역학과 분석역학으로 구분된다. 기술역학은 인구집단에서 건강, 질병 현상을 시간적, 지역적, 인적 변수 별로 기술하여 건강, 질병 빈도 차이를 일으키는 요인이 무엇인지에 대한 가설을 생성하는 역학 연구이다. 우리지역에서 어떤 감염병이 언제, 어디서, 누구에게 더 많이 발생하는지 기술하고 그 차이를 일으키는 요인에 대한 가설을 생성하는 것을 주요 목적으로 하는 감염병 감시분석보고서는 기술역학적 연구의 성격을 가진다.

질병관리청의 「역학조사관 교육수료 안내서」에서는 감시데이터에 대한 과학적 분석을 통해 감염병 예방 및 방역정책 수립의 합리적 근거자료를 생산하는 역량을 강화하는 것을 감시분석보고서 작성의 목적으로 제시하고 있다. 또한 감염병 감시데이터 분석을 통한 객관적 현황정보 생산, 감염병 감시데이터를 수집·분석하여 해당 질병의 지역별·시기별 역학적 특성 기술, 감염병의 주요 위험요인 분석 및 공중보건 측면에서의 함의 도출을 감시분석 보고서의 주요 내용으로 제안하고 있다.

2. 감염병 감시분석보고서 작성 틀

1) 과학 논문 및 보고서 등 과학적 글쓰기의 이상적 형태

과학적 논문 및 보고서 등 과학적 글쓰기의 이상적 형태는 IMRAD (Introduction, Methods, Results and Discussion) 구조이다(그림 9-3-1). 서론(Introduction), 방법(Methods), 결과(Results), 고찰(Discussion)의 네 부분으로 구성되며 서론은 고찰과, 방법은 결과와 대칭적인 형태를 가진다. 서론은 광범위한 내용에서 좁은 내용으로 고찰 좁은 범위에서 넓은 범위로 서술되는 경우가 많으므로 전

체적으로 모래시계와 같은 구조를 가진다. 서론에는 어떤 문제를 왜 연구하게 되었는지 서술하게 되며 연구를 이해하는 데 필요한 배경지식을 포함한다. 방법에는 이 연구의 방법을 최대한 자세히 서술해야 하는데, 다른 사람이 이 연구를 재현할 수 있을 정도로 자세히 기술해야 한다. 결과에는 이 연구를 통해 발견한 결과를 표와 그림으로 제시하고 글로 서술한다. 고찰에는 이 연구의 중요한 결과를 이전 연구 결과에 비춰 함의를 찾고, 연구의 약점과 강점, 향후 연구가 더 필요한 주제에 대해 서술할 수 있다.

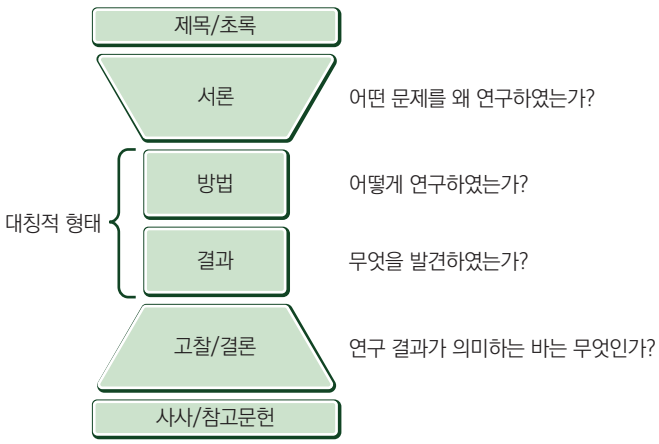


그림 9-3-1. 과학적 글쓰기의 이상적 형태

2) 감염병 감시분석보고서의 양식

감염병 감시분석보고서 역시 IMRAD 구조를 기본으로 한다. 질병관리청의 「역학조사관 교육수료 안내서」에 예시로 제시된 양식을 보면 제목, 서론, 대상 및 방법, 분석 결과, 결론 및 고찰의 순으로 기술하도록 하고 있으며, 각 부분의 세부 제목은 자유롭게 구성할 수 있다.

2 감염병 감시분석보고서의 자료원과 데이터 정리

감염병 감시분석보고서에서 활용할 수 있는 주요한 자료원은 전수감시 또는 표본감시 대상이 되는 감염병 신고자료와 사례조사서이다. 분석 대상 감염병을 선정하고 분석 기간을 설정하여 다운로드한다. 데이터가 제대로 다운로드 되었는지 확인하고 신고자료와 사례조사서를 하나의 데이터로 붙여서 준비하여야 한다. 중복된 데이터가 있는 경우 정리하고, 각 변수가 어떤 정보를 담고 있는지 하나씩 확인하

고, 날짜 변수를 연월일 단위로 분리, 연령을 그룹화하는 등 본격적인 분석에 들어가기 전에 분석 방법에서 계획한 분석을 수행하기 위한 사전 준비를 해야 한다.

감시분석보고서를 작성하기 위해서는 신고자료, 사례조사서 외 발생률 산출을 위한 연령별 인구 데이터가 필요하다. 통계청 국가통계포털(KOSIS) 홈페이지에서 다운로드하여 이용할 수 있다. 또한 해당 감염병의 우리 지역 발생률을 전국 발생률, 시도별 발생률과 비교할 필요가 있는 경우 질병관리청 감염병포털에서 발생률 자료를 다운로드하여 사용할 수 있다.

어떤 자료를 이용하더라도 이용한 모든 데이터를 언제, 어떻게 다운로드하였는지 방법에 상세히 기술하여야 하며, 중복 데이터 정리 변수 그룹화 등 분석을 위해 데이터를 정리한 내용도 방법에 상세히 기술하는 것이 좋다.

3 변수의 속성에 따른 표와 그래프의 작성과 서술

감염병 감시분석보고서는 그 특성상 결과가 보고서의 상당 부분을 차지하게 된다. 결과에는 분석 결과를 표와 그래프로 제시하고 표와 그래프에서 짚어야 할 내용이나 표와 그래프로 제시되지 않은 내용을 글로 풀어서 서술한다. 표는 많은 정보를 담을 수 있는 장점이 있으나 직관적인 해석과 비교가 어렵다는 단점이 있다. 반면 그래프는 정보가 생략될 수 있다는 단점이 있으나 직관적이며 시각적 비교가 가능하다는 장점이 있다. 따라서 정보를 많이 담을 수 있는 표가 보고서에서 더 많이 사용되지만 핵심적으로 제시하고 싶은 중요한 결과는 그래프로 제시하는 것이 좋다.

앞서 과학적 글쓰기의 이상적 형태에서 방법과 결과는 대칭적 형태로 기술되어야 한다고 하였다. 결과에서 제시되는 표와 그래프 등의 분석 방법은 방법 부분에 제시되어 있어야 한다. 예를 들어 발생률은 어떻게 산출하였는지, 변수의 종류별로 빈도와 분율은 어떻게 제시하였는지, 어떤 종류의 그래프를 어떤 변수를 이용하여 그렸는지 등이 기술되어야 한다.

표와 그래프를 작성하기 위해서는 변수의 종류(범주형, 연속형)에 따라 적절한 요약법을 선택해야 하며, 사전에 종이에 스케치를 해보고 표와 그래프를 디자인해 보는 것이 좋다.

1. 변수의 속성에 따른 요약 방법

변수는 통계학적으로 크게 범주형(categorical) 변수와 연속형(continuous) 변수로 나뉜다. 범주형 변

수는 남/녀, 행정구역, 결혼여부, 직업, 연령 범주(10, 20대 등) 등과 같이 범주로 구분하여 측정된 변수로 질적변수라고 하기도 한다. 연속형 변수는 길이, 몸무게, 온도 등 수치로 측정되어 덧셈이나 뺄셈 등의 연산이 가능한 변수로 양적변수라고 하기도 한다.

또한, 역학에서는 변수가 담고 있는 정보에 따라 시간적, 인적, 지역적 변수로 구분하기도 하는데 기술역학적 연구에서는 반드시 이 세 종류의 변수를 포함하도록 해야 한다. 감시분석보고서에서 주로 사용하는 시간적 변수는 연도, 월, 주차(week)가 있으며 이것은 날짜(일별로 측정된 연속형 변수)를 범주화 한 것이므로 범주형 변수로 구분할 수 있으나 경우에 따라 연속형 변수로 간주하여 분석하기도 한다. 또한 진단에서 신고까지 걸린 시간도 시간적 변수에 속하며 이것은 연속형 변수이다. 인적 특성을 담고 있는 변수로 성별, 연령 범주, 직업 등이 있으며 모두 범주형 변수이다. 연령이 범주화되지 않고 만 나이로 분석에 이용한다면 연속형 변수이다. 공간적 변수는 행정구역, 농촌/도시, 건물의 층 등이 있으며 모두 범주형 변수이다. 이와 같이 기술역학 연구에서 사용하는 많은 변수는 범주형 변수의 속성을 갖는다.

변수가 범주형인지 연속형인지 구분하는 것은 매우 중요한데, 이 속성에 따라 변수의 요약 방법이 다르기 때문이다. 범주형 변수를 요약하는 방법은 빈도표, 막대그래프, 원그래프 등이 있다. 빈도표에는 각 범주에 해당하는 대상자의 수를 제시한다. 단순히 수만 제시하지 않고 백분율(%)을 같이 제시하거나 인구 수를 분모로 계산한 발생률을 같이 제시할 수 있다. 원그래프를 사용할 때는 3차원 원그래프는 사용하지 않는 것이 좋은데 위치에 따라 크기가 왜곡될 수 있기 때문이다. 연속형 변수는 평균과 표준편차, 히스토그램 등으로 요약할 수 있다. 이와는 별개로 공간적 특성을 기술하는 특별한 방법으로 지도를 이용하는 방법이 있다. 지도위에 대상의 위치를 직접 표시하는 점지도가 있고, 발생률이 높은 지역은 짙은 색으로 낮은 지역은 옅은 색으로 표시하는 단계구분 지도가 있다. 단계구분 지도는 감시분석보고서 작성에도 많이 이용되는데 통계청에서 제공하는 SGIS 홈페이지에서 간단한 지도를 그려볼 수 있다.

2. 교차표

두 개 이상의 범주형 변수를 동시에 분석에 사용하여 교차표(cross tabulation)를 작성할 수 있다. 교차표란 두 가지 혹은 그 이상의 속성에 따라 분류된 관측 값으로부터 얻어지는 도수로 구성되는 데이터로 정의된다. 표 9-3-1의 (A)는 성별과 연령 그룹이라는 두 개의 범주형 변수로 교차표를 작성한 것이다. 이 교차표는 확장이 가능한데, 예를 들어 성별과 직업이라는 두 개의 범주형 변수를 이용해 교차표를 만들어 (A)의 아래로 붙일 수 있다. 마찬가지로 성별과 읍면동으로 교차표를 만들어 아래에 붙이는 방식으로 확장이 가능하다. 확장된 교차표는 감시분석보고서에서 자주 활용된다.

교차표에서는 세 종류의 백분율을 사용할 수 있다. 표 9-3-1의 (B)는 열 백분율, (C)는 행 백분율, (D)는 전체 백분율을 보여주고 있다. 세 종류의 백분율 중 저자가 보여주고자 하는 것이 무엇인지에 따라 하나를 선택하여 보여주면 된다. (B)는 남자와 여자의 연령 분포를 비교하고자 할 때 (C)는 연령 그룹에 따른 성별 분포를 비교하고자 할 때 (D)는 전체 중 해당 칸의 성별과 연령에 해당하는 사람이 몇 %인지 보여주고자 할 때 사용할 수 있으며 그에 따라 표의 제목을 작성한다.

3. 표와 그래프의 서술

표와 그래프를 작성하면 해당 표와 그래프의 주요 내용을 글로 서술해야 한다. 표와 그래프를 독자에게 읽어주는 것이다. 해당 표와 그래프에 담긴 정보 중 어떤 내용을 읽어줄지 먼저 정해야 한다. 시간에 따른 발생률의 변화를 보여주는 그래프라면 언제 발생률이 가장 높았는지, 발생률이 커지고 있는지, 작아지고 있는지 등을 서술할 수 있다. 표 9-3-1의 (B)의 경우 전체 환자의 연령 분포, 성별에 따른 연령 분포의 비교, 어느 연령이 가장 많은 분율을 차지하는지 혹은 가장 적은 분율을 차지하는지, 남자와 여자 중 어느 그룹이 더 많은지 등을 서술할 수 있다. 표와 그래프의 서술 방법에는 정답이 있는 것은 아니다. 저자가 무엇을 보여주려고 작성한 것인지, 독자에게 꼭 보여주고 싶은 중요한 발견은 없는지 검토하여 서술한다.

표 9-3-1. 교차표와 백분율의 종류

(A)

(단위: 명)

		전체	성별	
			남	여
전체		206	101	105
연령그룹	0~9세	2	0	2
	10~19세	8	4	4
	20~29세	20	7	13
	30~39세	41	19	22
	40~49세	54	30	24
	50세 이상	81	41	40

(B) 열 백분율

		전체		성별			
				남		여	
		N	%	N	%	N	%
전체		206	100.0	101	100.0	105	100.0
연령그룹	0~9세	2	1.0	0	0.0	2	1.9
	10~19세	8	3.9	4	4.0	4	3.8
	20~29세	20	9.7	7	6.9	13	12.4
	30~39세	41	19.9	19	18.0	22	21.0
	40~49세	54	26.2	30	29.7	24	22.9
	50세 이상	81	39.3	41	40.6	40	38.1

(C) 행 백분율

		전체		성별			
				남		여	
		N	%	N	%	N	%
전체		206	100.0	101	49.0	105	51.0
연령그룹	0~9세	2	100.0	0	0.0	2	100.0
	10~19세	8	100.0	4	50.0	4	50.0
	20~29세	20	100.0	7	35.0	13	65.0
	30~39세	41	100.0	19	46.3	22	53.7
	40~49세	54	100.0	30	55.6	24	44.4
	50세 이상	81	100.0	41	50.6	40	49.4

(D) 전체 백분율

		전체		성별			
				남		여	
		N	%	N	%	N	%
전체		206	100.0	101	49.0	105	51.0
연령그룹	0~9세	2	1.0	0	0.0	2	100.0
	10~19세	8	3.9	4	50.0	4	50.0
	20~29세	20	9.7	7	35.0	13	65.0
	30~39세	41	19.9	19	46.3	22	53.7

		전체		성별			
				남		여	
		N	%	N	%	N	%
연령그룹	40~49세	54	26.2	30	55.6	24	44.4
	50세 이상	81	39.3	41	50.6	40	49.4

4 보고서 작성 시 유의 사항과 수정대비표의 작성

1. 보고서 작성 시 유의 사항

질병관리청의 「역학조사관 교육수료 안내서」에 보고서 작성 시 유의 사항이 제시되어 있다. 이것은 역학조사보고서에도 공통적으로 적용되는 것으로 반드시 확인하고 준수하도록 한다. 이러한 규칙들은 독자의 입장에서 빠른 시간에 보고서를 읽고 내용을 파악하는데 도움을 준다. 특히 표와 그림 표기방법, 보고서 말미에 제시하는 참고문헌 양식을 준수하지 않는 경우가 종종 있으므로 주의 깊게 살펴야 한다.

2. 수정대비표의 작성

보고서의 심사결과는 승인, 조건부 승인, 재심사 가능, 재심사 불가의 네 종류로 통보된다. 이 중 조건부 승인이나 재심사 가능의 경우 수정대비표를 작성하여 제출하여야 한다. 수정대비표는 심사자가 지적한 내용에 대한 답변을 일목요연하게 정리하는 것으로 「역학조사관 교육수료 안내서」의 서식을 이용하여 작성한다. '심의 결과' 란에는 심사자의 지적 사항 하나하나에 각각 일련번호를 부여하여 지적사항을 그대로 복사하여 붙인다. 한 명의 심사자가 세 가지의 사항을 지적하였다면 세 개의 일련 번호가 부여된다. '수정 결과'에는 어떤 내용을 어떻게 수정했는지 답변한다. 수정 전과 후의 본문 내용은 수정 결과 란에 같이 쓸 수도 있다. 보고서에는 수정한 부분을 표시해 주어야 수정한 부분이 어디인지 심사자가 알 수 있으므로 반드시 수정한 부분을 붉은색 등으로 표시해야 한다.

5 맺음말

감염병 감시분석보고서 작성은 우리 지역의 감염병 현황을 파악하고 문제를 발견하기 위한 중요한 방법이다. 한 번의 보고서 작성으로 끝나는 것이 아니라 역학조사관이 지속적으로 수행하여야 하는 중요한 업무 중 하나이다. 감시분석보고서를 작성하고 심사를 받는 것은 역학조사관의 업무를 익히고 역량을 개발하는 과정이므로 공부하는 자세로 꼼꼼히 작성해 보도록 하자.

참고 문헌

1. 대한예방의학회. 예방의학과 공중보건학. 4판. 서울: 계축문화사; 2021.
2. 질병관리청. 역학조사관 교육수료 안내서. 2024.
3. Forero DA, Lopez-Leon S, Perry G. A brief guide to the science and art of writing manuscripts in biomedicine. J Transl Med. 2020;18(1):425.
4. Sollaci LB, Pereira MG. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. J Med Libr Assoc. 2004;92(3):364-7.
5. Thomas CG. Research Methodology and Scientific Writing. 2nd ed. Switzerland AG: Springer Nature; 2021.
6. Thomas Rahlf. Data Visualisation with R - 111 Examples. 2nd ed. Cham: Springer Nature; 2019.

제4절 역학조사 보고서의 작성

1 유행 및 역학의 개념

1. 유행

유행(epidemic, outbreak)은 특정 지역이나 집단, 또는 시간이 지나면서 질병이나 건강 관련 사건, 특정 상태가 예상보다 훨씬 더 많이 발생하는 상황을 말한다. 유행은 보통 일정 기간 동안 갑자기 증가하고, 특정 지역이나 장소에서 집중적으로 발생한다. 또, 어떤 인구 집단에서 발병률이 높아지는 경향이 있다. 유행과 관련된 용어는 엔데믹(endemic)과 팬데믹(pandemic)이 있다. 엔데믹은 특정 지역이나 집단에서 지속적으로 발생하는 질병을 뜻하고, 팬데믹은 전 세계적으로 퍼진 질병을 말한다.

2. 역학

역학(epidemiology)은 질병이나 건강 상태가 어떻게 분포되는지, 그 원인이 무엇인지 연구하는 학문이다. 이런 연구를 통해 질병의 원인이나 발생 패턴을 알아내고, 예방 방법도 찾을 수 있다. 역학의 주요 목적은 질병의 원인을 찾고, 질병이 발생하는 패턴을 파악하며, 예방할 수 있는 방법을 제시하는 것이다. 또, 보건 정책을 세우는 데 필요한 근거도 제공한다.

역학에서 중요한 개념은 세 가지이다. 첫 번째는 질병 분포로, 질병이 언제, 어디서, 누구에게 발생하는지를 분석하는 것이다. 두 번째는 결정 요인인데, 질병에 영향을 미치는 요인들은 생물학적 요인(병원체나 유전자), 환경적 요인(기후나 위생), 행동적 요인(생활 습관)이 있다. 세 번째는 질병 발생의 3요소인데, 병원체(질병을 일으키는 세균이나 바이러스 등), 숙주(병원체에 감염될 수 있는 사람이나 동물), 환경(병원체와 숙주가 상호작용하는 외부 조건)이 중요하다.

역학의 연구 방법은 여러 가지가 있다. 기술역학은 질병의 발생 패턴을 분석하는 방법으로, 유병률이나 발생률을 계산하는 것이 포함된다. 분석 역학은 질병의 원인과 그 연관성을 분석하는 방법으로, 환

자-대조군연구나 코호트연구가 있다. 실험역학은 실험을 통해 원인과 결과를 검증하는 방법이고, 임상시험이 대표적인 예이다.

2 유행 역학조사 단계

1. 유행 확인

유행 확인 단계는 유행이 실제로 발생했는지 확인하고 단순한 자료 오류나 자연적인 변동인지, 아니면 실제 유행인지 평가를 목적으로 한다. 활동내용은 발병률을 비교하여 과거의 발생률과 현재 상황을 비교하고 유행 여부를 확인한다. 자료 신뢰성을 검토하여 진단 기준 변경이나 보고 체계 강화 등으로 인한 오류 가능성을 배제한다.

2. 사례정의 및 확인

사례정의 및 확인 단계는 질병의 특징을 기준으로 조사 대상 사례를 명확히 규정하는 것을 목적으로 한다. 활동내용은 사례정의를 설정하여 질병의 임상적, 시간적, 공간적 기준을 포함해 조사 대상자를 선정한다. 사례정의를 개발하여 의심, 가능성, 확정 사례의 기준을 설정한다. 병원 기록, 감시체계 데이터, 인터뷰 등을 통해 사례를 확인하고 사례정의에 부합하는 환자를 선정하여 데이터를 수집한다.

3. 기초 정보 수집

기초 정보 수집 단계는 질병의 특성과 유행의 전반적 상황을 이해하기 위한 과정이다. 활동내용은 환자의 인구학적 정보(나이, 성별, 직업 등)와 임상적 정보(증상, 발병 시기, 치료 결과)를 수집한다. 환경적 요인을 평가하여 조사 지역의 위생 상태와 환경 오염 여부를 파악하고 위험 요인을 파악하여 감염 가능 경로나 노출 원인을 탐색한다.

4. 기술 역학 분석

기술 역학 분석 단계는 질병의 발생 양상을 시간, 장소, 사람의 관점에서 분석하는 것을 목적으로 한다. 활동내용은 시간적 분석을 통해 발생 곡선(Epidemic Curve)을 작성하고, 발병 시점을 기준으로 유행의 시작, 정점, 종료 시기를 파악한다. 공간적 분석을 통해 지도나 그래프를 이용해 질병 발생의 지리적 분포를 확인한다. 뿐만 아니라 인구학적 분석을 통해 발병률이 특정 인구 집단(나이, 성별 등)에서 높게 나타나는지 평가한다.

5. 가설 설정

가설 설정 단계는 유행의 원인, 전파 경로, 주요 위험 요인에 대한 가설을 수립하는 것을 목적으로 한다. 활동내용은 기술 역학 분석 결과를 바탕으로 가능한 원인을 도출한다. 예를 들어, 특정 음식을 섭취한 사람들이 설사 증상을 보였다면, 특정 음식이 감염의 주요 원인일 가능성이 있다.

6. 분석 역학 조사

분석 역학 조사 단계는 가설을 검증하고 원인과 위험 요인을 규명한다. 활동내용은 환자-대조군연구를 통해 질병에 걸린 사람과 걸리지 않은 사람을 비교하여 위험 요인을 분석한다. 코호트연구를 통해 특정 집단을 대상으로 노출 여부와 발병 여부를 추적하여 관련성을 평가하며 통계 분석을 통해 위험도, 오즈비(OR), 상대 위험도(RR) 등을 계산한다.

7. 환경 조사

환경 조사 단계는 병원체나 유행 원인이 되는 환경적 요인을 확인하는 것을 목적으로 한다. 활동내용은 음식, 물, 공기, 토양 등 환경 검체를 수집하고 실험실 검사를 진행한다. 또한 감염원이 존재하는 환경 조건을 분석한다.

8. 예방 및 통제 조치

예방 및 통제 조치 단계는 유행을 멈추고 재발을 방지하는 것을 목적으로 한다. 활동내용은 즉각적인

조치로 감염원을 제거하고, 환자를 격리하거나 예방 접종을 실시한다. 위험 요인을 통제하여 오염된 물 공급을 중단하고, 안전한 음식 공급을 보장한다. 지역 사회에 예방 수칙을 홍보하고 정보를 제공한다.

9. 결과 해석 및 보고

결과 해석 및 보고 단계는 조사 결과를 정리하고 정책 및 대책에 반영하는 것을 목적으로 한다. 활동내용은 조사 결과를 논리적으로 해석하여 결론을 도출하여 보건 당국 및 관련 기관에 보고서를 제출하고, 결과를 학술적으로 발표하거나 출판한다.

10. 후속 조치 및 모니터링

후속조치 및 모니터링 단계는 예방 조치가 효과적으로 시행되었는지 확인하는 것을 목적으로 한다. 활동내용은 유행 종료 후에도 지속적인 감시 체계를 운영하는 것이다. 이 과정에서 예방 조치의 성과를 평가하고, 필요시 추가 조치를 취한다.

3 역학조사 보고서의 개요

1. 목적

역학조사 보고서는 감염병 유행의 원인 규명 과정과 방역 조치 결과를 기술하고 역학조사 과정에서 얻은 교훈을 공유한다. 이를 통해 유행이 어떻게 발생했는지, 무엇을 했고, 무엇을 발견했는지 설명한다. 또한 조사 결과에서 도출된 결론을 검증한다. 여기서는 조사 방법, 방역 조치, 윤리적 이슈 등을 확인하고 유사 사례 대응 역량을 향상시키고, 역학조사 정보를 공유하는 것을 활성화한다. 또한 유행 발생과 질병 자연사의 변화를 기록한다. 예를들어 비병원성이나 초기 병원성 단계에서는 환경 개선과 건강 증진이 중요하고 불현성 감염에서 현성 감염으로 이어지는 과정에서는 조기 진단과 치료가 중요하다. 회복 단계에서는 재활이 필요하다. 이러한 과정을 통해 조사 프로토콜을 개선한다.

2. 계획 수립

역학조사 준비 단계에서 보고서 작성을 고려하는 것은 중요하다. 최초 상황을 평가한 후, 조사 목적과 방법을 수립하고 역학조사를 위한 적절한 과학적 지식을 검토한다. 이를 위해 문헌을 검토하고 유사 사례를 확인하며, 질병의 원인, 전파 방식, 위험 요인을 파악한다. 현장 조사를 통해 위험 요인을 재평가하고 필요시 설문지를 준비하고, 분석 계획을 수립한다. 최종적으로 자료를 정리하여 라인 리스팅을 한다.

3. 작성 및 제출 양식

역학조사 보고서는 제목, 개요, 초록, 서론, 방법, 결과, 결론 및 고찰, 참고문헌을 기본으로 하여 작성한다(표 9-4-1).

표 9-4-1. 역학조사 보고서 작성 양식

기본양식	주요 내용	비고
제목	분석대상감염병, 지역 및 기간 등 명시(필요시 연구방법 포함)	Title
개요(표)	발생 지역, 규모, 원인병원체, 추정감염경로 등 기술	Summary
초록	배경 및 목적, 조사방법, 조사결과, 결론 순으로 작성	Abstract
서론	유행발생 배경 및 사건 설명	Introduction (Background and rationale, objective)
본론	조사대상, 조사방법(사례정의, 조사설계, 설문조사, 실험실검사, 분석방법)	Method
결과	발병률, 유행곡선, 주요 증상, 위험요인 분석, 실험실 검사, 환경 조사 결과	Result
결론 및 고찰	추정감염원(원인병원체), 감염병관리, 조사의 제한점	Conclusion and Discussion
참고문헌	저자, 제목(책, article 등), 학술지명(또는 학회명), 연도;권(호):페이지 * 밴쿠버 양식(Vancouver style)	Reference

4 역학조사 보고서 작성방법

1. 제목

제목에는 분석 대상인 감염병명, 지역, 발병 장소, 기간을 명시하고, 필요시 연구 방법도 포함하여 유행 사례를 간략하고 정확하게 반영한다. 예를 들어, "경상북도 ○○시 ○○요양병원 코로나19 집단 발생 역학조사 보고서"처럼 작성한다. 연구 방법을 표기할 경우, 예시로 "Leukemia incidence among workers in the shoe and boot manufacturing industry: a case-control study"와 같이 기술할 수 있다.

2. 개요

개요(표)에는 발생 신고 일시, 추정 위험 노출 일시, 현장 역학 조사 일시, 최초 사례 발생 일시, 발생 지역, 위험 노출 인구, 조사 설계, 잠복기, 발생 규모, 원인 병원체, 장소 특성, 추정 감염 경로, 인구 집단 특성, 조치 사항 등을 개요 표 양식에 맞게 기입한다(표 9-4-2).

표 9-4-2. 개요(표) 작성 양식(예시)

발생신고일시	2025년 00월 00일(월) 00:00	추정위험 노출일시	2025년 00월 00일(일)
현장역학조사 일시	2025년 00월 00일(목)~2025년 00월 00일(금)	최초사례 발생일시	2025년 00월 00일(일)
발생지역	충북 청주시	위험노출인구	약 000명
조사설계	사례군조사, 후향적코호트, 환자/대조군 등	잠복기	00일~00일(평균 00일)
발생규모	발병률(확진 000명, 노출 000명)	원인병원체	SARS-CoV-2
장소특성	요양병원	추정 감염경로	병원 내 비말전파
인구집단특성	입원환자 및 의료진 등	조치사항	코호트격리, 표면소독

3. 초록

초록은 독자가 이해할 수 있도록 보고서 전체를 요약하고, 모호하지 않으며 간략하고 명료하게 작성한다. 배경 및 목적, 조사 방법, 조사 결과, 결론의 구조로 작성하며, 배경 및 목적에는 연구 수행 동기와 목적, 해결하고자 하는 문제를 설명한다. 조사 방법에는 문제 해결을 위한 접근 방식을 기술하고, 조사

결과에는 연구 결과를 요약하며, 결론에는 연구 결과가 미치는 영향과 해결책을 제시한다. 이때 인용문, 표, 그림, 출처, 약어 등은 사용하지 않는다.

4. 서론

서론에서는 감염병 유행의 배경과 목적을 설명한다. 배경에서는 유행의 원인 사건을 명확히 제시하고, 유행이 최초 보고된 방법, 조사팀 구성 및 조사 시작 시기 등 초기에 취한 조치를 설명한다. 또한, 유행 판단의 근거인 사람, 장소, 시간에 대한 상황을 설명하며, 과학적 근거를 바탕으로 감염병의 특징을 설명한다. 목적에서는 유행 조사 목적에 병원체, 감염원, 전파 경로, 위험 요인 등을 포함하여 명확하게 설명한다. 조사자의 조사 이유와 잠재적인 가설을 설정하며, 조사 목적이 향후 유사 사례 대응을 위한 참고 자료로 활용되므로 잘못된 예시는 피해야 한다.

5. 본론

본론에서는 재현이 가능하도록 절차와 도구를 상세히 기술한다. 조사 대상자 구분 기준, 조사 방법, 사례정의, 조사 설계 등을 설명하며, 설문 조사 방법과 실험실 검사 방법도 포함된다. 또한, 분석 방법과 관련 통계 프로그램 및 연관성 측정 도구도 상세히 기술한다.

6. 결과

결과에서는 발견한 내용을 기술하고 연구 방법을 반영한 결과를 도출한다. 발병률, 유행 곡선, 주요 증상, 위험 요인 분석, 실험실 검사 결과 등을 포함하며, 시간, 공간, 사람의 특성을 고려한 유행 곡선을 제시한다. 실험실 검사 결과와 환경 조사도 포함된다.

7. 결론 및 고찰

결론 및 고찰에서는 연구 목적에 부합하는 주요 결과를 요약하고, 연구 결과에 대한 해석과 논의를 이어간다. 유행 조사 및 대응의 강점과 약점을 논의하고, 연구 결과의 공중보건학적 의미를 도출한다. 추가 발생 억제를 위한 방역 조치와 중단기 예방을 위한 중재안을 제시하며, 연구의 한계를 인식하고 논의한다.

8. 참고문헌

참고문헌은 본문에서 언급한 순서대로 번호를 기재하며, 최근 보고된 연구를 활용한다. 참고문헌은 밴쿠버 양식을 사용하며, 저자명은 성-중간 이름-이름 순으로 표기한다.

9. 작성내용 최종점검

작성 내용은 자가 점검표와 역학조사 심사 기준을 활용하여 최종 점검한다(표 9-4-3, 4).

표 9-4-3. 역학조사보고서 자가점검표

점수	구분	내용
5점	탁월	구성요소를 포함하고 있고, 안내된 내용에 부합하며, 질적으로 우수함
4점	충분	구성요소를 포함하고 있고, 안내된 내용에 부합함
3점	보통	구성요소를 포함하고 있으나, 안내된 내용에 완전히 부합하지는 않음
2점	불충분	구성요소를 포함하고 있으나, 결함 또는 질적으로 부족한 점이 있음
1점	누락	구성요소가 누락됨
N/A	해당없음	수행한 연구·조사와 관계없음

구분	내용	1	2	3	4	5	N/A
서론 (Introduction)	1. 유행의 원인이 된 사건/유행이 최초 보고된 방법/팀이 어떻게 구성되었는지/조사 시작 시기 등 설명						
	2. 사람, 장소, 시간에 대한 상황 설명						
	3. 병원체 또는 질병 설명(원인, 임상 특성, 풍토성 또는 계절성)						
	4. 유행사례 조사 이유에 대한 기술 또는 가설 설정						
	5. 유행 역학조사의 목적을 명확하게 설명						
방법 (Methods)	1. 연구 설계 및 시간성(후향적, 전향적, 양방향) 제시						
	2. 사례정의 및 조사 대상자 선정 기준 포함						
	3. 사례 발굴 및 조사대상자 모집방법 설명						
	4. 조사대사자에 대한 데이터 수집방법 기술						
	5. 편향을 제거하는 방법 설명						
	6. 조사 대상자 유형 설명/조사 대상자 규모 계산						
	7. 분석에 사용된 통계적 방법과 소프트웨어를 포함한 데이터 분석 절차 설명						
	8. 임상 및 환경 검체채취/실험실 검사 방법 설명						

구분	내용	1	2	3	4	5	N/A
방법 (Methods)	9. 인간 피험자를 보호하는 데 사용되는 프로세스를 포함(기밀성, 위험, 사전 동의)_IRB						
결과 (Results)	1. (필요시) 설문 응답률과 함께 참가자 수를 명시						
	2. 장소와 공간적 분포 등을 포함, 명확하고 간결한 정보 제공 (불필요한 미사여구 지양)						
	3. 유행 곡선 또는 시간에 따른 발생 양상 제시						
	4. 연령, 성별 및 기타 관련 변수별 발병률 제공						
	5. (필요시) 정밀도 측정(신뢰 구간)을 사용하여 효과 크기를 보여주고 교란변수에 따라 조정						
고찰 (Discussion)	1. 결과에 기술한 내용을 바탕으로 핵심 결론 도출						
	2. 참고문헌으로 결론을 뒷받침(결론을 지지하지 않는 데이터에 대해서도 참고문헌을 활용하여 설명)						
	3. 데이터, 연구 설계, 역학조사 장소, 가능한 편향 등에 대한 제한점을 기술						
	4. 추가 발생 억제를 위한 즉각적인 방역조치 보고						
	5. 중재 연구 결과의 외적 타당성에 대한 고찰 기술(예: 결과를 다른 집단이나 장소에 어느 정도까지 일반화할 수 있는지)						
	6. 장기 중재의 타당성과 지속 가능성에 대한 고찰 기술						
	7. 향후 유사한 유행사례을 예방하는 방법 또는 현재 감염병 유행을 해결하기 위해 필요한 추가 연구에 대한 제안						
구조 (Structure)	1. 제목(유행사례를 간략하고 정확하게 반영)						
	2. 저자 및 기여자(기관 포함)						
	3. IMRD 형식 사용(서론, 방법, 결과, 결론 및 고찰)						
	4. 연구방법을 반영한 결과 도출						
	5. 결과 부분에 연구방법이 작성되지 않도록 함						
	6. 고찰이나 결론에서 결과가 반복되지 않도록 함						
	7. 제안사항은 결과를 기반으로 작성						
참고문헌 (References)	1. 참고문헌은 본문에 언급한 순서대로 번호 기재						
	2. 보고서 내용에 적합한 참고문헌의 수						
	3. 보고서 내용과 관련이 있으며, 최근 보고된 연구를 활용						
	4. 참고문헌 작성은 Vancouver Style을 사용						
일반양식 및 명확성 (General form and clarity)	1. 적절한 구두점과 문법 사용						
	2. 적절한 맞춤법 사용						
	3. 보고서는 독창성을 지녀야 하며, 표절 금지						
	4. 글은 단락 형식이며, 글머리 기호 사용 금지						
	5. 약어에 대한 설명 제공						

표 9-4-4. 역학조사보고서 심사기준

심사기준	세부요건
1. 연구주제 및 문제의 명료성	• 역학조사의 목적을 명확하게 기술
	• 역학조사 출동을 결정한 타당한 근거 제시
	• 감염병 발생을 인지한 경위와 당시 상황을 구체적으로 제시
2. 연구방법의 타당성	• 초기상황 및 현장상황 파악을 통한 적절한 조사방법 선택
	• 주요 증상 및 위험요인에 대해 객관적으로 기술
	• 전체집단 구성과 조사대상자의 선정기준을 정확히 명시
3. 관련 문헌과 자료분석의 충실성	• 연구에 착수하기 위한 선행 데이터 또는 연구결과 제시
	• 결과의 해석이 현재의 과학적 지식과 일치
	• 보고서 내용과 관련 있으며, 최근 보고된 연구결과 활용
	• 조사대상 감염병의 현재 상황 요약·제시
	• 시간, 장소, 상황 등을 모두 고려한 사례정의 도출
4. 논리전개의 적절성	• 각 파트별 내용의 적절한 배치 및 조화로운 논리적 흐름 유지
	• 조사방법을 반영한 결과 도출
	• 결과에 근거하여 종합한 결론 제시(분석 데이터 반복 X)
5. 보고서의 질적 수월성	• 역학조사 사례를 간략하고 정확하게 반영한 제목 제시
	• 간결하고 명료한 문장으로 작성
	• 역학적 주요 개념의 정확한 적용
	• 효과적인 이해와 비교가 가능한 수치자료 배치
	• 산출된 통계 데이터에 대한 객관적 설명
	• 사람, 시간, 장소를 포함한 표·그림 제목 제시
	• 결과를 강조할 수 있는 최소한의 수치를 활용한 표·그림 작성
6. 연구결과의 정책적 기여도	• 공중보건 조치의 근거가 될 수 있을 만큼 타당하고 신뢰할 수 있는 결론 제시
7. 연구윤리의 충실성	• 자료 분석 및 해석에 개인적인 가치 또는 주관 개입 배제
	• 연구대상 개인 또는 단체 익명성 확보
	• 참고문헌은 본문에 언급한 순서대로 번호 기재
	• 보고서 내용에 적합한 참고문헌의 수
	• Vancouver 형식에 따라 작성된 참고문헌

참고 문헌

1. 질병관리청. 역학조사관 교육수료 안내서. 2024.
2. CDC. Principles of Epidemiology in Public Health Practice. 3rd ed. 2006.
3. ECDC. Guidelines for writing outbreak investigation reports. 2019.

4. Public Health Agency of Canada. Outbreak reporting guide. Can Commun Dis Rep. 2015;41(4):73–5.
5. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Epidemiology. 2007;18(6):805–35.

제5절 초록 작성 및 학술 발표

1 초록 작성하기

1. 초록의 특성과 구조

초록(Abstract)[1]은 논문이나 보고서에서 제목과 함께 본문 앞에 위치하며 시각적으로 중요한 요소이다. 초록은 독자의 관심을 이끌어내야 하므로 간결하면서도 명확하고 구체적이며 매력적이어야 한다. 의학 관련 논문이나 보고서는 대체로 배경 및 목적, 방법, 결과 및 결론으로 구조화된 초록을 요구한다. 초록의 끝부분에는 키워드를 제시하는 경우가 많고, 키워드는 MeSH (Medical Subject Headings) 키워드를 활용하는 것이 좋다. 초록은 다음과 같은 특징을 갖춰야 한다.

- 구체적이고 명확할 것
- 편향되지 않고 정직할 것
- 간결하면서도 정밀할 것
- 독립적으로 이해 가능하도록 완결성을 갖출 것
- 학술적일 것

초록은 보통 독자가 논문의 전체 내용을 빠르게 개관하기 위해 가장 먼저 읽는 부분이지만, 작성은 마지막에 하는 경우가 많다. 초록은 일반적으로 짧게 작성되지만(보통 150~300단어), 학술 논문이나 학술대회 발표에서 매우 중요한 부분이다. 잘 작성된 초록은 독자가 본문을 읽고 싶어지도록 충분한 흥미를 유발할 수 있다. 초록은 연구 전체를 개괄적으로 요약하여 특히 그 결과와 학문적 기여를 강조하도록 작성되어야 하며, 독자가 논문을 읽기 전에 해당 논문이 자신에게 적합한지 판단할 수 있도록 도

1 "Abstract"라는 단어는 라틴어 abstractum에서 유래했으며, 긴 글을 압축한 형태를 의미한다.

와야 한다. 초록은 다음 질문에 대한 답변을 포함해야 한다.

- 이 논문은 무엇에 관한 것인가?
- 이 연구가 왜 중요한가?
- 어떻게 연구를 수행했는가?
- 무엇을 발견했는가?
- 이러한 발견이 왜 중요한가?

각 학문 분야와 학술지, 학술대회에는 특정한 양식과 분량의 초록을 요구하는 경우가 있으므로 이를 확인하여 초록을 작성하도록 한다.

2. 좋은 초록 작성하기

초록을 작성하는 순서는 다음과 같다. 요구되는 초록의 구조와 단어 수를 확인한 후 논문이나 보고서를 다시 읽어 개요를 파악한다. 연구 배경과 목적, 방법, 결과, 토론, 결론의 각 부분을 읽고 주요 내용을 1~3 문장으로 요약한다. 요약된 문장을 다시 읽으면서 중요한 내용이 누락되지 않았는지 확인한다. 단어 수를 확인하고, 불필요한 단어를 삭제하거나 문장을 더 간결하게 재작성하여 길이를 줄인다. 문장의 흐름과 표현이 매끄럽도록 다듬는다.

좋은 초록은 다음과 같은 특징을 가지므로 이를 참고하여 작성하도록 한다.

- 하나의 잘 구성된 단락으로, 논리적이고 간결하며 독립적으로 정보를 전달할 수 있도록 한다.
- 주요 학술적 요소인 배경, 목적, 방법, 결과, 결론을 모두 포함하도록 한다.
- 본문에 포함되지 않은 내용이 초록에 포함되지 않도록 한다.
- 쉽고 명확하게 작성하여, 특정 학문 분야의 독자뿐만 아니라 더 넓은 독자층이 이해할 수 있도록 한다.
- 일반적으로 참고 문헌은 포함하지 않는다.

2 학술 발표하기

역학조사관은 학술대회나 세미나, 워크숍 등 다양한 형태의 학술모임에서 학술적 발표를 해야 하는 경우가 있다. 학술 발표에 있어 핵심 메시지를 선정하고 발표의 구조를 잡고, 자료를 준비하여 실제로 효과적으로 전달되도록 발표하는 방법을 개괄적으로 안내하고자 한다.

1. 핵심 메시지와 발표의 범위 정하기

훌륭한 발표에는 명확한 메시지가 필요하다. 이 핵심 메시지는 발표의 중심이 되어 전체 구조를 결정한다. 명확한 메시지를 설정하면 단순히 정보를 전달하는 것에서 벗어나 보다 효과적으로 내용을 전달할 수 있다. 핵심 메시지의 깊이와 너비는 발표 시간에 따라 결정될 수 있다. 발표 시간을 엄수하는 것은 학술 발표에서 매우 중요한 부분이므로 발표 시간은 핵심 메시지와 발표의 범위를 정할 때 매우 중요하게 고려해야 할 요소이다. 전달하고 싶은 정보의 양에 집착하기보다, 주어진 시간 내에 몇 가지 포인트와 얼마나 많은 세부사항을 전달할 수 있는지 고려해야 한다. 예를 들어, 10분짜리 발표에서는 일반적으로 두세 가지 복잡하지 않은 핵심 메시지만 다룰 수 있고 배경과 방법 등은 깊이 있게 다루기 어렵다.

2. 발표내용 구성하기

핵심 메시지와 발표 내용의 범위를 정리한 후, 다음 단계는 발표 내용을 체계적으로 구성하는 것이다. 다른 형태의 학술 글쓰기와 마찬가지로, 발표 역시 세 부분으로 나눌 수 있다. 도입부에서는 발표의 목적과 구조를 설명하고, 본문에서는 주요 포인트를 다루며, 결론에서는 발표를 요약하고 그 중요성을 강조하는 것이 좋다.

1) 도입부

청중의 관심과 흥미를 끌기 위해 이야기나 최근 동향에 대한 논평을 활용하여 중요한 질문, 문제, 혹은 논쟁를 제기할 수 있다. 또는 간단한 맥락과 배경 설명으로 발표를 시작한 후, 핵심 메시지가 다루는 문제나 논쟁을 간결하게 설명할 수도 있다. 두 경우 모두, 도입부의 다음 단계는 발표의 목적이나 핵심 메시지를 명확히 제시하는 것이다. 다음과 같은 표현을 사용할 수 있다.

"오늘 저는 매우 논란이 되는 주제에 대해 이야기하려 합니다..."
"이 질문은 ...를 이해하는 데 핵심적입니다."
"저는 ...라는 주장을 펼치고자 합니다."
필요한 경우, 다음과 같은 표현으로 발표의 범위를 제한할 수 있다.
"여러 이론이 있지만, 이 발표는 두 가지에만 초점을 맞춥니다..."
"공공 부문이 아닌 민간 부문에만 초점을 맞춥니다..."
"정책 형성이 아니라 구현 과정에 대해 논의하겠습니다..."
"제 논점은 세 가지 주요 포인트를 기반으로 합니다. 첫째로... 둘째는... 그리고 마지막으로..."

2) 본문

본문에서는 핵심 메시지를 뒷받침하는 내용을 다룬다. 각 주요 포인트를 명확하고 논리적인 순서로 설명하며, 각 포인트가 서로와 핵심 메시지에 어떻게 관련되는지 연결 지어야 한다. 다음과 같은 표현을 활용할 수 있다.

"다음 포인트로 넘어가 보겠습니다..."
"또 다른 중요한 고려 사항은..."
"앞서 살펴본 내용을 기반으로, 이제 ...에 대해 이야기하겠습니다."

필요한 개념과 용어는 사용하기 전에 반드시 정의하고 설명해야 하며, 예시를 제시하면 좀 더 효과적인 설명이 가능하다.

3) 결론

발표가 끝났음을 알리는 표현의 예로 들면 다음과 같다.

"결론적으로..."
"마지막으로 말씀드리자면..."

발표에서 다룬 주요 포인트를 요약하며, 주제의 중요성을 상기시키고 발표의 목적을 다시 언급한다. 발표에서 제시한 목표가 어떻게 달성되었는지 설명하며 청중에게 감사를 표하고, 질문이나 의견을 받도록 한다.

3. 시각화 도구 활용하기

발표에 시각 자료를 반드시 사용해야 하는 것은 아니지만, 이를 잘 활용하면 여러 가지 이점이 있다. 시각적으로 표현하면 청중이 내용을 더 잘 이해할 수 있으며, 관심을 유도하고 단순히 구두로 전달하는 것보다 더 강한 인상을 남길 수 있다. 그러나 시각 자료가 발표 내용 자체를 대체해서는 안 된다. 발표 구조와 주요 포인트를 정한 이후에 슬라이드나 배포 자료를 준비한다.

파워포인트(PowerPoint) 등을 이용한 슬라이드는 발표 그 자체가 아니라 발표를 보조하는 역할을 한다. 슬라이드에는 몇 가지 핵심 키워드, 이미지, 또는 데이터를 포함해 발표 내용을 강조하고 데 초점을 맞추어야 한다. 슬라이드에 포함되지 않은 세부 내용, 예시, 추가 정보는 발표자가 구체적으로 설명하도록 해야 한다. 파워포인트 슬라이드를 작성할 때 다음을 유념하도록 한다.

- 단순하게 유지하기: 복잡하거나 산만한 슬라이드는 피해야 한다. 슬라이드는 발표자의 말을 보조하는 도구이지, "발표의 주인공"이 되어서는 안 된다.
- 시각적 매체 활용하기: 텍스트 대신 이미지를 활용한다. 간결한 표와 차트는 정량적 정보를 효과적으로 전달할 수 있고 사진이나 그래픽은 생생한 내용을 전달할 수 있게 해준다.
- 일관성과 간결성 유지하기: 단순한 폰트를 사용하고, 슬라이드 전체에서 일관성을 유지한다. 텍스트는 맨 뒷자리에서도 읽을 수 있을 만큼 충분히 크고 명확해야 한다. 색상을 사용할 때는 밝은 조명 아래에서도 잘 보이는 조합을 선택하고, 필요 이상으로 많은 정보를 한 슬라이드에 넣지 않도록 한다.
- 출처를 명확히 밝히기: 다른 사람의 아이디어, 이미지, 데이터 등을 인용한 경우, 반드시 적절히 출처를 밝혀야 한다. 말로 직접 언급할 수도 있으나 슬라이드에 출처를 표기하는 것이 좋다. 발표 마지막에 참고 문헌 리스트 슬라이드를 포함하도록 한다.

4. 구두 발표하기

마지막 단계는 구두로 발표를 하는 것이다. 발표할 때는 미리 준비한 '대본'을 그대로 읽는 것은 지양해야 하며, 슬라이드나 노트를 참고하여 이야기하듯 말하는 것이 좋다. 이야기하듯 전달하는 방식은 읽는 방식보다 훨씬 덜 단조롭고, 청중과의 시선 교환 및 상호작용을 더욱 원활하게 한다. 구두 발표에서는 글로 작성된 논문이나 보고서보다 약간 비공식적이고 구어체적인 언어 사용이 허용된다. 또한, 발표 중에 자신의 입장을 요약하거나 앞서 했던 설명으로 돌아가는 방식(backtracking)도 내용 전달에 효과

적이다.

발표에 대해 긴장하거나 불안감을 느끼는 것은 지극히 자연스러운 일이다. 기술 문제로 인해 발표가 잘못될 수 있다는 두려움, 실수를 하거나 질문에 답하지 못할까 하는 두려움은 공적인 발표를 하는 모든 사람이 공통적으로 겪는 감정이다. 약간의 긴장감은 오히려 좋은 영향을 줄 수 있으며 발표를 성공적으로 이끌 수도 있다. 긴장감을 다루고 통제하는 가장 좋은 전략은 철저히 준비하는 것임을 잊지 말자.

연습은 훌륭한 발표를 하기 위해 필수적이다. 연습은 발표를 개선하고 자신감을 심어주며, 어려운 단어나 외국어 발음에 익숙해지도록 해준다. 또한 시각 자료와 발표 개요를 익히는 데도 유용하다. 가장 중요한 것은 시간 제한을 준수할 수 있도록 해준다는 점이다. 연습 방법은 다양하다. 예를 들어, 동료 학생, 친구 또는 가족에게 피드백을 요청하거나, 컴퓨터나 거울 앞에서 혼자 연습할 수 있다. 자신을 녹화하고 다시 보는 것도 훌륭한 전략이다. 이를 통해 잘하고 있는 부분과 개선이 필요한 부분을 명확히 알 수 있다.

실전에서는 깊게 숨을 들이쉬고, 차분하고 여유롭게 시작한다. 준비가 잘 되어 있다는 점을 기억하고 자신감을 가지도록 한다. 발표를 할 때는 다음 사항에 주의를 기울이자.

- 충분히 큰 목소리로 모든 청중이 들을 수 있도록 말하기
- 천천히 말하며 핵심 포인트를 강조하기
- 명확하게 말하기: 억양은 중요하지 않지만, 청중이 내용을 이해할 수 있어야 함
- 청중을 바라보고 좋은 시선 교환을 유지하기: 노트만 보거나 한 곳만 바라보지 않도록 주의함
- 열정을 보여주기: 본인이 발표 내용에 흥미가 없다면, 다른 사람도 흥미를 느끼기 어려움
- 긍정적인 신체 언어: 등을 곧게 펴고 손을 자연스럽게 유지하며 청중의 주의를 산만하게 하는 행동(예: 서성거리기, 몸을 구부리기 등)은 피해야 함

발표 끝뿐만 아니라 발표 중에도 질문에 답변할 준비가 되어 있어야 한다. 청중이 질문을 하면, 정중하게 응답하고 긍정적으로 받아들이도록 한다. 질문은 청중이 발표 내용을 경청하고 흥미를 느끼고 있다는 증거이다. 또한, 청중과의 토론을 유도하기 위해 몇 가지 질문을 준비해 두는 것도 좋다. 만약 답할 수 없는 질문을 받는다면 당황하지 말고, 절대 얼버무리지 않는다. 대신, 질문의 중요성을 인정하고 솔직하게 모른다고 말한 뒤, 청중 중에 답변할 수 있는 사람이 있을지 물어보는 것도 한 방법이다. 연구 주제와 관련된 모든 내용을 완벽히 알 수는 없다.

5. 훌륭한 학술 발표를 위한 6가지 팁

1) 파워포인트를 신중하게 활용하기

이미지는 강력한 도구로 기억과 학습에 도움을 준다. 메시지를 효과적으로 전달할 수 있는 이미지를 찾아 사용하는 것은 효과적이다. 파워포인트는 훌륭한 도구지만, 효과적으로 사용할 때만 그렇다. 일반적으로 시각 자료를 많이 사용하고, 텍스트는 최소화해야 한다. 24포인트 이하의 폰트를 사용하지 말고, 슬라이드 내용을 읽지 않도록 한다.

2) 학술 발표에는 공식이 있다. 이를 따르라

발표 전문가가 되면 공식을 벗어날 수 있겠지만, 초보자라면 이를 따르는 것이 좋다. 학문 분야에 따라 약간 다를 수 있지만, 의학과 보건학 분야는 대체로 다음과 같은 공식을 따른다.

- 도입/개요/흥미 유발
- 이론적 배경/연구 질문
- 방법론/사례 선택
- 데이터/결과 논의
- 결론

3) 청중은 당신의 연구에 대해 듣고 싶어 한다. 그것을 말하라

발표에서 가장 흔한 실수 중 하나는 청중이 이미 알고 있는 정보를 반복하는 것이다. 이는 기존 문헌이나 배경 정보에 너무 많은 시간을 할애하는 경우 생길 수 있는 실수이다. 발표 시간의 절반가량은 이 연구의 방법과 결과에 할애하는 것이 좋다. 학술 발표에서 청중이 기대하는 것은 기존 연구 요약이 아니라, 새롭고 흥미로운 연구 결과이다.

4) 연습, 연습, 연습하라

발표 전에는 반드시 전체 발표를 연습해야 한다. 집에서 혼자, 또는 동료나 가족 앞에서 반복적으로 연습해야 한다. 연습을 통해 발표가 시간 제한 내에 들어가는지 확인하고, 발표의 흐름을 더 자연스럽게 만들 수 있다. 연습은 많을수록 좋다.

5) 시간 제한을 준수하라

발표 시간이 10분이라면, 10분 분량의 자료만 준비한다. 7분만 주어진다면, 반드시 그 안에 끝내야 한다.

6) 발표를 읽지 말라

누군가가 발표문을 읽는 것을 듣고 몰입하기는 어렵다. 반드시 읽어야 한다면, 청중이 당신이 읽고 있다는 사실을 모를 정도로 자연스럽게 읽도록 한다. 이를 위해서는 대화체로 작성하고, 여러 차례 연습하며, 감정과 확신을 담아 톤을 조절하면서 읽는 연습이 필요하다.

참고 문헌

1. Australian National University [Internet]. Academic Skills [Cited 2024 Dec 30]. Available from: https://www.anu.edu.au/students/academic-skills.
2. Forero DA, Lopez-Leon S, Perry G. A brief guide to the science and art of writing manuscripts in biomedicine. J Transl Med. 2020;18(1):425.
3. The University of Adelaide [Internet]. Writing an Abstract: Writing Centre Learning Guide [Cited 2024 Dec 30]. Available from: https://www.adelaide.edu.au/writingcentre/ua/media/26/learningguide-writinganabstract.pdf.
4. Thomas CG. Research methodology and scientific writing. 2nd ed. Cham, Switzerland: Springer Nature; 2021.
5. Wiley [Internet]. 6 tips for giving a fabulous academic presentation [Cited 2024 Dec 30]. Available from: https://www.wiley.com/enus/network/publishing/research-publishing/writing-and-conducting-research/6-tips-for-giving-a-fabulous-academic-presentation.

제6절 STROBE와 ORION 가이드라인의 이해: 역학조사 결과의 논문 작성 및 연구결과의 공표

역학조사관으로서 역학조사 보고서를 학술대회에 발표하거나 논문으로 작성하는 것은 중요하고 매우 가치 있는 일이다. 왜냐하면 자신이 알아낸 보건학적 문제나 해결방안을 여러 사람과 공유함으로써 더 많은 사람에게 혜택을 줄 수 있기 때문이다. 또한 역학조사 보고서나 감시자료 분석보고서는 형식적인 측면에서 학술논문으로 발전시킬 수 있는 형태로 작성되기 때문에 추가적인 노력으로 단순히 보고서로만 그치지 않고 중요한 내용은 학술논문으로 발전시켜서 더 많은 사람에게 알릴 필요가 있다.

그렇지만 역학조사결과를 논문으로 재작성한다는 것은 현실적으로 쉽지만은 않다. 이것은 논문작성에 대한 교육과 훈련이 필요하다. 여기서 소개하는 STROBE와 ORION은 각각 관찰역학연구와 의료관련감염유행조사의 논문작성 가이드라인이다. 이 가이드라인을 소개하는 것은, 만약 이 가이드라인의 내용을 알고 있다면 역학조사를 수행하거나 감시자료를 분석할 때 대상을 누구로 하고, 무엇을 조사하고, 어떤 자료를 어떻게 정리해야 논문으로 작성할 수 있을지에 대한 내용을 알 수 있기 때문이다. 이러한 개념이 없이 역학조사나 감시자료를 분석하였을 때 나중에 논문으로 작성하고자 하면, 논문의 제반 구성요소에서 미처 조사하지 못한 부분이나 빠뜨린 부분들이 있어서 논문작성시 어려움이 있을 수 있기 때문이다. 이 내용을 잘 이해하고 있다면 역학조사나 감시자료분석 시 논문을 염두에 두고 진행할 수 있다.

참고로 역학조사와 감시자료분석한 내용과 관련해서 가장 관련 있는 국내 학술지는 한국역학회지(Epidemiology&Health)와 대한예방의학회지(Journal of Preventive Medicine&Public Health), 질병관리청에서 발간하는 Osong Public Health&Research Perspective가 있고, 감염병 역학에 관한 것은 대한감염학회와 화학요법학회가 공동으로 발간하는 Infection & Chemotherapy가 대표적이다. 이들 학술지를 포함하여 해외에서 발간되는 기타 학술지들에 따라서는 각각 요구하는 논문 구성이나 표현은 조금씩 다를 수 있지만, 일반적인 관찰역학논문 작성은 이들 가이드라인을 따른다.

1 관찰역학연구의 논문작성 가이드: STROBE

연구형태에 따라 연구를 논문화할 때 필요한 가이드라인이 만들어져 있다. 가장 처음 만들어진 것은 1996년 발표된 CONSORT (Consolidated Standards of Reporting Trials)이고 여기서는 무작위대조임상시험(RCT) 논문의 질을 높이기 위한 가이드라인을 담고 있다. CONSORT는, 이후에 만들어지는 각종 논문 가이드라인의 기본 틀을 제시한 점에서 매우 의미가 있다. CONSORT 이후 공중보건분야에서 확률화 할당이 되지 않은 인터벤션 연구의 가이드라인으로 TREND statement (Transparent Reporting of Evaluations of Nonrandomized Designs)가 발표되었고, 관찰역학연구논문의 가이드라인을 만들기 위한 작업은 2003년부터 시작되었다. STROBE initiative (Strengthening the Reporting of Observational studies in Epidemiology)는 단면조사연구, 환자-대조군연구, 코호트연구 등 각각의 관찰역학연구에 대해서 논문에 필요한 가이드라인으로, 몇 차례의 개정을 거쳐서 2007년 각종 학술지에 결과가 발표되었다.

역학조사는 기본적으로 환자-대조군 혹은 후향성코호트로 정리될 수 있다. 또한 역학조사관이 수행하는 다양한 연구는 관찰역학연구에 포함되므로 이를 여기에 간략히 소개한다.

STROBE 가이드라인의 구성과 설명은 표 9-6-1과 같다. 이 각각의 항목에 대한 설명(배경, 의의, 작성요령)등 자세한 내용은 뒤에 제시한 참고문헌을 참고한다.

표 9-6-1. STROBE 지침: 관찰역학연구에서 포함되어야할 요소와 체크리스트

	번호	권고내용
Title&Abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/ rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection

	번호	권고내용
Participants	6	(*a*) *Cohort study*–Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*–Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*–Give the eligibility criteria, and the sources and methods of selection of participants
		(*b*) *Cohort study*–For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*–For matched studies, give matching criteria and the number of controls per case
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) *Cohort study*–If applicable, explain how loss to follow-up was addressed *Case-control study*–If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*–If applicable, describe analytical methods taking account of sampling strategy
		(*e*) Describe any sensitivity analyses
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram

	번호	권고내용
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
		(b) Indicate number of participants with missing data for each variable of interest
		(c) *Cohort study*–Summarise follow-up time (eg, average and total amount)
Outcome data	15*	*Cohort study*–Report numbers of outcome events or summary measures over time
		Case-control study–Report numbers in each exposure category, or summary measures of exposure
		Cross-sectional study–Report numbers of outcome events or summary measures
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		(*b*) Report category boundaries when continuous variables were categorized
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done–eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

* Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

2 역학조사의 논문작성 가이드: ORION

ORION은 병원감염(원내유행)에서 필요한 ourtbreak study를 기술할 때의 가이드라인이라 지역사회 유행조사 연구의 틀과 조금 다른 부분이 있지만, 기본적인 항목은 같다(표 9-6-2). 자세한 내용은 참고

문헌을 활용한다.

표 9-6-2. Checklist of items to include when reporting an outbreak or intervention study of a nosocomial origin

	Item number&Descriptor
Title & Abstract	1. Description of paper as outbreak report or intervention study. Design of intervention study (e.g. ITS with or without control group, cross-over study). Brief description of intervention and main outcomes.
Introduction	
Background	2. Scientific and/or local clinical background and rationale. Description of organism as epidemic, endemic or epidemic becoming endemic.
Type of paper	3. Description of paper as intervention study or an outbreak report. If an outbreak report, report the number of outbreaks.
Dates	4. Start and finish dates of the study or report.
Objectives	5. Objectives for outbreak reports. Hypotheses for intervention studies.
Methods	
Design	6. Study design. Use of EPOC classification29 recommended (CBA, or ITS). Whether study was retrospective, prospective or ambidirectional. Whether decision to report or intervene was prompted by any outcome data. Whether study was formally implemented with predefined protocol and endpoints.
Participants	7. Number of patients admitted during the study or outbreak. Summaries of distributions of age and lengths of stays. If possible, proportion admitted from other wards, hospitals, nursing homes or from abroad. Where relevant, potential risk factors for acquiring the organism. Eligibility criteria for study. Case definitions for outbreak report.
Setting	8. Description of the unit, ward or hospital and, if a hospital, the units included. Number of beds, the presence and staffing levels of an infection control team.
Interventions	9. Definition of phases by major change in specific infection control practice (with start and stop dates). A summary table is strongly recommended with precise details of interventions, how and when administered in each phase.
Culturing and typing	10. Details of culture media, use of selective antibiotics and local and/or reference typing. Where relevant, details of environmental sampling.

	Item number&Descriptor
Infection-related outcomes	11. Clearly defined primary and secondary outcomes (e.g. incidence of infection, colonization, bacteraemia) at regular time intervals (e.g. daily, weekly, monthly) rather than as totals for each phase, with at least three data points per phase29 and, for many two phase studies, 12 or more monthly data points per phase.20 Denominators (e.g. numbers of admissions or discharges, patient bed days). If possible, prevalence of organism and incidence of colonization on admission at same time intervals. Criteria for infection, colonization on admission and directly attributable mortality. All cause mortality. For short studies or outbreak reports, use of charts with duration patient stay and dates organism detected may be useful (see text).
Economic outcomes	12. If a formal economic study was done, definition of outcomes to be reported, description of resources used in interventions, with costs broken down to basic units, stating important assumptions.
Potential threats to internal validity	13. Which potential confounders were considered, recorded or adjusted for (e.g. changes in length of stay, case mix, bed occupancy, staffing levels, hand-hygiene compliance, antibiotic use, strain type, processing of isolates, seasonality). Description of measures to avoid bias including blinding and standardization of outcome assessment and provision of care.
Sample size	14. Details of power calculations, where appropriate.
Statistical methods	15. Description of statistical methods to compare groups or phases. Methods for any subgroup or adjusted analyses, distinguishing between planned and unplanned (exploratory) analysis. Unless outcomes are independent, statistical approaches able to account for dependencies in the outcome data should be used, adjusting, where necessary, for potential confounders. For outbreak reports statistical analysis may be inappropriate.
Results	
Recruitment	16. For relevant designs, such as cross-over studies,30 or where there are exclusions of groups of patients, the dates defining the periods of recruitment and follow-up, with a flow diagram describing participant flow in each phase.
Outcomes and estimation	17. For the main outcomes, the estimated effect size and its precision (usually using confidence intervals). A graphical summary of the outcome data is often appropriate for dependent data (such as most time series).
Ancillary analyses	18. Any subgroup analysis should be reported and it should be stated whether or not it was planned (i.e. specified in the protocol) and adjusted for possible confounders.
Harms	19. Pre-specified categories of adverse events and occurrences of these in each intervention group. This might include drug side effects, crude or disease-specific mortality in antibiotic policy studies or opportunity costs in isolation studies.

	Item number&Descriptor
Discussion	
Interpretation	20. For intervention studies an assessment of evidence for/against hypotheses, accounting for potential threats to validity of inference including regression to mean effects and reporting bias. For outbreak reports, consider clinical significance of observations and hypotheses generated to explain them.
Generalizability	21. External validity of the findings of the intervention study, i.e. to what degree can results be expected to generalize to different target populations or settings. Feasibility of maintaining an intervention long term.
Overall evidence	22. General interpretation of results in context of current evidence.

3 역학조사결과 해석의 주의점

역학조사결과는 통계 분석을 통하여 산출된 결과를 적절히 해석해야 한다. 해석은 연구자 자신이 연구의 결과를 이해하고 보다 넓은 맥락에서의 연구 결과의 의미를 찾는 과정이라고 할 수 있다. 해석은 연구에서 발견된 사항을 설명하는 데에서 시작한다. 여기서 중요한 사항은 발견된 사항 중 자신의 결론과 잘 맞는 부분과 잘 맞지 않는 부분을 일목요연하게 정리하는 것이다. 역학조사는 실험실 연구와 달리 통제된 환경이 아닌, 관찰연구이기 때문에 실제 관찰과 해석에서 여러 방향이 다 가능한 경우가 많다. 연구자는 무의식적으로 자신의 결론과 맞는 부분만 강조하고 맞지 않는 것은 무시하려는 경향이 생길 수 있으며, 자칫 잘못된 해석으로 일관할 수가 있다.

〈Tip 7〉 결과 해석상 주의해야 할 점

① 무리한 해석
② 표본오차의 무시
③ 응답범주의 수가 너무 적을 때 생기는 문제
④ 실제적인 의미가 없이 통계적인 의미가 있는 관계에 집착하지 말 것
⑤ 인과관계에 대한 지나친 언급: 연구자가 의도하지 않더라도 논문을 읽는 독자나 언론은 이를 확대 해석하는 경향이 있으므로, 인과관계에 대한 언급에서는 명료하게 부합되는 것과 그렇지 않은 것을 정리한다.
⑥ 태도와 행동의 혼동
⑦ 전후상황 무시

해석에서 특히 관심을 쏟아야 할 사항은 연구 방법론에 관한 것이다. 자신의 연구의 방법론의 약점을 정확히 파악하고 이것이 연구결과에 미치는 영향에 대하여 충분히 고려하여야 한다. 연구자는 결과의 해석과정에서 자신의 연구결과에 대하여 비판적인 안목을 잊지 말아야 한다.

4 유행 역학조사 논문에 대한 제언

질병 유행조사 시 역학조사보고서를 작성하는 것이 일반적이고, 역학조사보고서의 구성와 내용이 현장 중심의 보고서 형식으로 되어 있어서, 이를 논문화하는 데는 참고문헌의 고찰 등 일정한 수고가 들어가게 된다. 그러나 역학조사보고서의 기본 구성도 기술역학적 방법을 통한 가설설정과 분석역학적 방법을 통한 가설검정이 모두 이루어지고, 이를 논리적으로 구성하면 좋은 논문이 된다. 따라서 역학조사를 할 때 이를 처음부터 논문화하기 위한 마음자세를 가지고 임하는 것이 필요하며, 역학조사는 유행시기만 가능하며 유행이 지나면 유행시기에 관찰해야 할 것들을 놓칠 수 있으므로 현장에서 조사를 진행할 때부터 철저히 수행하는 것이 좋다.

미국의 EIS conference에서는 EIS들이 조사한 역학조사 사례를 중심으로 학술대회가 구성되며, 이 중 일부는 MMWR을 비롯한 학술지에 실린다. 관심을 갖고 review해 보면 외국에서 학술지에 실리는 것과 우리나라에서 역학조사 내용이 논문으로 실리는 것에는 약간의 차이가 있음을 알 수 있다.

예를 들어, 잔치집에서 발생한 살모넬라 식중독 사례를 논문화하는 경우를 생각해 보자. 환자에게서 살모넬라균이 검출되었고, 기존 역학조사 방법을 따라 그대로 조사를 마쳤다면, 이 연구를 논문으로 받아들일 수 있을지에 대한 판단은 사독자에 따라 달라질 수 있다. 살모넬라 식중독은 비교적 흔하고, 이미 자연사 및 역학적 특성이 잘 알려져 있으며, 조사 결과도 기존 지식을 크게 벗어나지 않는 경우가 많기 때문이다. 이처럼 게재를 긍정적으로 보는 입장과 그렇지 않은 입장은 각각 타당한 논리를 갖고 있다.

하지만 모든 사례 보고가 곧 논문이 된다고 생각하기보다는, 해당 역학조사를 통해 새롭게 보고할 만한 가치가 있는 요소를 발견하고자 하는 자세가 중요하다.

1. 역학조사 논문의 내용에 관련된 제언

논문의 생명은 독창성이다. 그런데 유행조사 논문은 일종의 사례연구로서 임상논문의 사례보고 내지는 사례군 보고와 유사한 성격을 갖는다는 점에 유의해야 한다. 만약 흔한 사례에 대해서 이미 잘 알려진

사실이 기술된 사례보고라면 어떤 사독자라도 이를 학술지에 싣는 것을 거절할 것이다. 살모넬라 식중독과 같이 이미 잘 알려진 질병유행이 비록 잘 기술되었더라도 대상이나, 접근방법, 역학적 특성이 기존에 알려진 것과 똑같고 그동안 잘 알려지지 않았던 새로운 결과 제시가 없다면 학술지에 굳이 지면을 할애할 이유가 없다.

질병유행조사 논문으로서 독창성 인정은 최종적으로 그 분야의 전문가(사독자)가 독자적으로 판단할 문제이나 다음과 같은 기준을 제시할 수 있으며 이러한 기준을 중심으로 종합적으로 판단할 문제라고 생각된다.

① 역학조사에서 새로운 감염원(기존에 보고되지 않은), 감염경로의 발견 및 보고
② 새로운 조사 방법론을 사용(통상적인 조사방법 이외에 추가된 것 등)하여 그 결과를 평가한 내용 포함
③ 유행질병에 대한 새로운 역학적 사실 보고
통상보다 광범위한 대상의 유행조사, 여러 개의 유행을 종합 분석하여 하나의 유행사례로서 볼 수 없었던 새로운 역학적인 특성의 기술 등
④ 기존에 유행하지 않았던 집단에서의 유행, 통상적으로 알려진 기술역학적 특성(인적, 시간적, 장소적 특성)과 다르게 유행한 사례
⑤ 새롭거나 비교적 드문 질병의 유행조사에 대한 보고
⑥ 환자관리나 항생제내성패턴 등 기존 역학조사보고사례에서 잘 알려지지 않은 새로운 내용 포함
⑦ 외국 문헌에는 잘 정리되어 있으나, 국내에서 보고가 거의 없는 내용 포함(예를 들면 세균성 이질의 2차 발병률은 국내에서 발표된 것이 적으므로 의미가 있다. 혹은 장티푸스에서 만성보균자의 비율계산이 이루어진 경우, 또 쯔쯔가무시증은 흔하고 이미 잘 알려져 있는 질병이나 이에 대한 국내 역학조사결과는 매우 적으므로 이에 해당된다고 볼 수 있는 등)
⑧ 위에서 언급한, 대상, 질병, 전파경로, 역학적 특성 등 이외에 새로운 사실이 추가되어 학문적 독창성이 인정된다고 판정되는 경우

2. 역학조사 논문의 구성(형식)에 관련된 제언

유행조사논문도 일반 연구논문의 형식을 그대로 따르나, 그 구성에 있어서 다음의 내용을 포함하는 것이 바람직하다.

① 유행 인지 경위
② 유행조사 방법
③ 유행에 대한 기술역학적 접근결과 기술
④ 유행에 대한 분석역학적 결과 기술(환자-대조군 혹은 후향성 코호트 설계방법에 적합해야 하며, 기타 하이브리드 접근방법을 사용할 수 있다)
⑤ intervention 방법과 그 결과
⑥ 실험실 결과(미생물학적, 분자생물학적, 화학적, 물리적 분석 등)

①~③번까지만 포함된 경우는 사례보고에 더 가깝다고 할 수 있다. 우리나라 역학조사의 발전을 위해서는 ④번과 ⑤번의 결과가 제시된 논문에 우선순위를 주는 것이 바람직하다고 판단한다.

역학조사는 훌륭한 논문의 소재이고, 역학조사보고서를 썼다면 논문의 반 이상은 쓴 것이나 다름없다. 염두에 두어야 할 것은 역학조사에 임할 때부터 논문화하는 것을 구상하고, 새로운 것을 찾거나 시도하는 자세가 필요하다는 것이다.

이제는 역학조사 자료를 표준화하여 데이터베이스화하는 것도 필요하다. 이 자료들은 하나씩만 보면 대단한 것이 아닐지 몰라도 우리나라 전체로 모인다면 보이지 않던 새로운 역학적 특성이 나타날 가능성이 매우 높다. 이러한 것으로 눈을 돌려보자.

5 맺으며

역학 조사를 연구설계에 맞게 수행하고 분석하고 어떤 연구논문까지 쓴다는 것은 매우 어려운 일임에 틀림없다. 다만 여러분의 역학조사관 교육 과정 중에서 만나게 될 많은 문제점들의 해결과정을 찾는 노력 중에 획득하는 지식들을 논문화하는 것을 적극 권장한다. 여러분의 건투를 빈다.

참고 문헌

1. Stone SP, Cooper BS, Kibbler CC, et al. The ORION statement: guidelines for transparent reporting of outbreak reports and intervention studies of nosocomial infection. Lancet Infect Dis. 2007;7(4):282-8.
2. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epide-

miology (STROBE): explanation and elaboration. Epidemiology. 2007;18(6):805-35.

Textbook For
Korea Field Epidemiology
Training Program





부록

찾아보기







ㄱ

가구접촉조사 연구 ····· 219
가변수 ····· 084
가설검정 ····· 003, 043
가축전염병 ····· 337
간접효과 ····· 214
간헐 반복노출형 ····· 143
감시 ····· 260
감시분석보고서 ····· 389
감염관리 ····· 347
감염 사슬 ····· 255
감염예방 및 관리 ····· 254
감염원 ····· 142
감염재생산수 ····· 164
개별짝짓기 ····· 156
검역 ····· 284
검역방법 ····· 296
검출바이어스 ····· 365
결과변수 ····· 007
결정계수 ····· 081
결측 ····· 372
계절성 ····· 112, 320, 352
계통오차 ····· 185
공기주의 ····· 257
공중보건 ····· 225
공통노출일 ····· 142
과산포계수 ····· 127
관찰단절 ····· 216
관찰역학연구 ····· 417
교란변수 ····· 368
교란인자 ····· 153
교란효과 ····· 153
교차표 ····· 392
구간추정 ····· 040
귀무가설 ····· 044
근거수준 ····· 386
기술역학 ····· 397
기술통계 ····· 002
기여 요인 ····· 179
기준범주 ····· 085
기준점 ····· 189
기초감염재생산수 ····· 164
기후변화 ····· 351
깔대기그림 ····· 383

ㄴ

내적 타당도 ····· 367
누적합 알고리듬 ····· 128

ㄷ

다변량분석 ····· 157, 160
다제내성균 ····· 342, 346
다제내성녹농균 감염증 ····· 249
다제내성아시네토박터바우마니균 감염증 ·· 249
다중공선성 ····· 087
다중 로지스틱 회귀모형 ····· 100
다중위치 염기서열 타이핑 ····· 195, 208

다중유전자패널검사 ····· 197
다중회귀분석 ····· 079, 083
단면 데이터 ····· 008
단면 연구 ····· 367
단순 로지스틱 회귀모형 ····· 099
단순회귀분석 ····· 079
단일공동 노출형 ····· 143
단절적 시계열분석법 ····· 221
단측검정 ····· 044
대립가설 ····· 044
대응표본 t-검정 ····· 057, 093
대조군의 선택 ····· 217
더빈-왓슨 검정 ····· 113
뎅기열 ····· 356
독립변수 ····· 007
독립성 검정 ····· 071
동질성 검정 ····· 071
등간형 ····· 005
등분산 가정 ····· 054

ㄹ

로그정규분포 ····· 143, 149
로지스틱 회귀분석 ····· 096
리드-프로스트 모형 ····· 167

ㅁ

막대그래프 ····· 027
말단수리 ····· 196
말라리아 ····· 353, 354
매개체 감염 ····· 353
맥니마 검정 ····· 074
맨텔-핸젤 보정법 ····· 160
맨-휘트니 검정 ····· 090
메타게놈 분석 ····· 203
메타분석 ····· 378
메티실린내성황색포도알균 감염증 ····· 249
명목형 ····· 005
모비율 ····· 039
모수 ····· 002
모수적 방법 ····· 089
모집단 ····· 002, 003
무작위배정 ····· 212
무작위 오차 ····· 186
민감도 ····· 185, 187

ㅂ

반복요소 PCR ····· 195, 208
반응변수 ····· 007
반코마이신내성장알균 감염증 ····· 249
반코마이신내성황색포도알균 감염증 ····· 249
발병률 ····· 167
백신 거부 ····· 239
백신 소통 ····· 229
백신 소통의 핵심 원칙 ····· 240
백신 신뢰 ····· 229
백신의 수용 ····· 228
백신의 영향 ····· 213

백신 주저 ····· 230
백신 주저 개념 모델 ····· 236
백신 주저 모델 ····· 234
범주형 ····· 008, 064, 391
병원체 보유자 ····· 261
병원체의 생존 ····· 179
보정 ····· 153
부호 검정 ····· 090, 094
분산 ····· 023
분산계수 ····· 149
분산분석 ····· 060
분석 역학 ····· 399
분절화 ····· 196
불규칙 변동 ····· 112
불편추정량 ····· 023
비례위험 가정 ····· 108
비말주의 ····· 257
비모수적 방법 ····· 089
비선형성 ····· 353
비율형 ····· 005
비정상성 ····· 112, 118
비표본오차 ····· 004
빈도분석 ····· 019
빈도짝짓기 ····· 156

ㅅ

사례 검사-음성군 연구 ····· 218
사례-대조군연구 ····· 217
사례정의 ····· 217, 398
사례-코호트연구 ····· 220
사분위수 범위 ····· 024
산술평균 ····· 022
산점도 ····· 030
산포도 ····· 022
상관계수 ····· 076
상관분석 ····· 076
상대위험도 ····· 065
생물정보학 도구 ····· 200
생물정보학 파이프라인 ····· 200
생존분석 ····· 106
생존함수 ····· 107
샤피로-윌크 검정 ····· 048
샷건 메타게놈 시퀀싱 ····· 204
서열형 ····· 005
선택 바이어스 ····· 365
설명변수 ····· 007
소통 ····· 225, 226
소통 단계 ····· 241
소통 요소 ····· 242
수정대비표 ····· 395
스크리닝법 ····· 220
시계열 데이터 ····· 008
시계열분석 연구 ····· 221
시계열분해 ····· 120
시계열 자료 ····· 112
시차 ····· 113
신뢰구간 ····· 040, 041
신뢰도 ····· 185
실험변수 ····· 007

실험역학 ····· 398

ㅇ

안티 백서 ····· 239
야생생물법 ····· 337
양성 예측도 ····· 185, 187
양적 자료 ····· 008
양측검정 ····· 044
엑솜 시퀀싱 ····· 197
엔데믹 ····· 397
역추적 조사 ····· 176, 177
역학 ····· 397
역학연구 ····· 364
연구설계 ····· 364
연속형 ····· 391
연속형 확률변수 ····· 034
예방접종 ····· 228
예방접종 의향 ····· 230
예방접종 참여 ····· 228, 230
오염 ····· 179
오즈 ····· 067
오즈비 ····· 067
외적 타당도 ····· 367
우도 함수 ····· 097
운송수단 ····· 291
워크시트 ····· 010
원헬스 ····· 318, 323
위양성 ····· 185, 187
위음성 ····· 185, 187
위험비 ····· 108, 216
위험함수 ····· 107
윌콕슨 부호순위 검정 ····· 090, 094
윌콕슨 순위합검정 ····· 090
유병률 ····· 187
유의수준 ····· 045
유의확률 ····· 046
유전자 감시 ····· 295
유전체학적 접근 ····· 195
유행 ····· 126
유행곡선의 성장률 ····· 168
유행모수 ····· 164
유효감염재생산수 ····· 164
음성 예측도 ····· 185, 187
의료관련감염 ····· 248, 266
의료관련감염병 ····· 249, 260, 266
이동 윈도우 ····· 127
이동 평균 ····· 122
이산형 확률변수 ····· 034
이상치 ····· 127
인수공통감염병 ····· 323, 330
일반화우도비 알고리듬 ····· 129
일원분산분석 ····· 060
일치(백분)율 ····· 192
임상검체 ····· 261

ㅈ

자기상관성 ····· 112, 113
자연내성 ····· 342

잔차 분석 ····· 086
잡음 ····· 112
재연 ····· 179
전장유전체 단일염기다형성 분석법 ····· 206
전장유전체분석 ····· 200
전장유전체 염기서열분석 ····· 195, 197
전체효과 ····· 214
전파확률측정 연구 ····· 219
점추정 ····· 040
접촉주의 ····· 257
정규분포 ····· 035
정규성 검정 ····· 048
정밀도 ····· 185
정보바이어스 ····· 366
정확도 ····· 185, 187
제1종 오류 ····· 045
제2종 오류 ····· 045
조기이상치보고체계 ····· 128
종속변수 ····· 007
주기성 ····· 112
중도절단 ····· 106
중심극한정리 ····· 038
중심척도 ····· 022
중위수 ····· 022
증식 ····· 179
지속 노출형 ····· 143
직접효과 ····· 214
질적 자료 ····· 008, 064
짝짓기 ····· 156
쯔쯔가무시증 ····· 357

ㅊ

차세대염기서열분석 ····· 195
참조 유전체 ····· 199
처치변수 ····· 007
체계적문헌고찰 ····· 375
초록 ····· 408
최대우도추정 ····· 097
최빈값 ····· 022
최소제곱법 ····· 097
추론통계 ····· 002
추세 ····· 112
추정 ····· 040
출판바이어스 ····· 377
측정 척도 ····· 004
층화분석 ····· 157

ㅋ

카바페넴내성장내세균목 감염증 ····· 249
카이제곱 검정 ····· 070
카파통계량 ····· 192
코로나19 백신 ····· 231
코호트연구 ····· 215, 364
콕스 비례위험 모형 ····· 106, 107
콜모고로프-스미노프 검정 ····· 048
크루스칼-왈리스 검정 ····· 090, 095

ㅌ

타당도 ····· 185
통계량 ····· 003
통계적 추론 ····· 003
특이도 ····· 185, 187

ㅍ

파이차트 ····· 026
패널 데이터 ····· 008
패링턴 알고리듬 ····· 127
팬데믹 ····· 397
펄스필드겔전기영동 ····· 195, 208
평균세대기 ····· 164
평균 잠복기 ····· 142
표본 ····· 003
표본분포 ····· 037
표본오차 ····· 004
표준정규분포 ····· 036
표준주의 ····· 256
표준편차 ····· 023
풍토병 ····· 303
피벗테이블 ····· 020
피셔의 정확검정 ····· 073

ㅎ

합계효과 ····· 214
항생제 ····· 340
항생제내성 ····· 319
항생제 스튜어드십 ····· 347
해외여행 ····· 303
해외유입 감염병 ····· 305, 308
핵심 유전체 다중 유전자좌 서열 분석 ····· 206
혼합형 차트 ····· 033
확률 ····· 034
확률밀도함수 ····· 035
확률변수 ····· 034
확률분포 ····· 035
확률질량함수 ····· 035
확장전파형 ····· 143
환자 ····· 261
환자-대조군연구 ····· 217, 365
회귀분석 ····· 079
회귀진단 ····· 086
획득내성 ····· 342
효과 ····· 212
효과수정인자 ····· 154
효능 ····· 212
히스토그램 ····· 027

기호

-2 로그 가능도 ····· 102

번호

2차발병률 ····· 219
3세대 염기서열분석법 ····· 201

5C 모델 ····· 235
16S rRNA 유전자 ····· 204