creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

보건학 석사학위논문

Epidemiological Application of Cycle Threshold Values Under The COVID-19 Pandemic in Korea

대한민국의 코로나-19 팬데믹 하에서 Cycle Threshold Values의 역학적 활용

2023년 2월

서울대학교 보건대학원

보건학과 보건학전공

김 지 원

Epidemiological Application of Cycle Threshold Values Under The COVID-19 Pandemic in Korea

지도교수 황 승 식

이 논문을 보건학석사 학위논문으로 제출함
2022년 11월

서울대학교 보건대학원
보건학과 보건학전공
김 지 원

김지원의 보건학석사 학위논문을 인준함
2022년 12월

위 원 장 _____ 조 성 일 _____ (인)

부위원장 _____ 이 우 주 _____ (인)

위 원 _____ 황 승 식 _____ (인)

CONTENTS

1. INTRODUCTION.....1

1.1 Study Background ……. 1

1.2 Systematic review …….. 2

1.3 Study objective ……..…….…..……. 8

2. METHODS 9

2.1 Study design …….…….. 9

2.2 Data source …...…… 10

2.3 Sample size ……... 11

2.4 Statistical analysis …… 13

3. RESULTS.....16

3.1 Descriptive statistics for diagnostic test results ……... 16

3.2 Distribution patterns of Ct values in Korea ……. 20

3.3 Forecasting the number of newly confirmed cases based on the Ct values ……. 27

3.3.1. Static model: Simple Linear regression model …… 27

3.3.2. Dynamic model: Finite Distributed Lag Model (FDL Model) …… 34

4. DISCUSSION 44

4.1 Summary of results ….…..…… 44

4.2 Comparison with previous studies …...…… 45

4.3 Strengths and Limitations ……... 46
4.4 Public health implications ……... 47

5. CONCLUSION.....48

REFERENCES.....49

APPENDIX.....53

ABSTRACT.....82

LIST OF TABLES

Table 1. Search terms used in national/international database .. 5
Table 2. Comparison with previous forecast studies..... 7
Table 3. Descriptive statistics for diagnostic test results of samples 18
Table 4. Simple Linear regression results on the number of newly confirmed cases by lagged Ct values RdRp (by date, all periods) 29
Table 5. Simple Linear regression results on the number of newly confirmed cases by lagged Ct values RdRp (by date, 1st COVID-19 periods)..... 30
Table 6. Simple Linear regression results on the number of newly confirmed cases by lagged Ct values RdRp (by date, 2nd COVID-19 periods)..... 31
Table 7. Simple Linear regression results on the number of newly confirmed cases by lagged Ct values RdRp (by date, 3rd COVID-19 periods)..... 32
Table 8. Simple Linear regression results on the number of newly confirmed cases by lagged Ct values RdRp (by date, 4th COVID-19 periods)..... 33
Table 9. Finite Distribution lag model of the number of newly confirmed cases by Ct values RdRp (by date) 36
Table 10. T-test results of linear combinations of estimators 37
Table 11. Finite Distribution lag model of the number of newly

confirmed cases by Reproduction rate (by date)..... 40

LIST OF FIGURES

Figure 1. Flow diagram of literature and selection criteria..... 6
Figure 2. Flow diagram of sample selection 12
Figure 3. Violin plot of Ct values E, RdRp, and N 19
Figure 4. Time series distribution graph of Ct values RdRp and Reproduction rate (by week). 23
Figure 5. Time series distribution graph of Ct values RdRp and number of newly confirmed cases (by date). 26
Figure 6. Simple Linear regression graph on the number of newly confirmed cases by lagged Ct values RdRp (by date, all periods). 29
Figure 7. Simple Linear regression graph on the number of newly confirmed cases by lagged Ct values RdRp (by date, 1st COVID-19 periods). 30
Figure 8. Simple Linear regression graph on the number of newly confirmed cases by lagged Ct values RdRp (by date, 2nd COVID-19 periods). 31
Figure 9. Simple Linear regression graph on the number of newly confirmed cases by lagged Ct values RdRp (by date, 3rd COVID-19 periods). 32
Figure 10. Simple Linear regression graph on the number of newly confirmed cases by lagged Ct values RdRp (by date, 4th COVID-19 periods). 33
Figure 11. LPR of Ct values on newly confirmed cases 38
Figure 12. LPR of Rt on newly confirmed cases 41

Figure 13. LPR of Rt on newly confirmed cases in the time of Omicron variants spread..... 43
Figure 14. LPR of Ct values on newly confirmed cases in the time of Omicron variants spread..... 43

1. INTRODUCTION

1.1. Background information

코로나-19(COVID-19) 진단은 SARS-CoV-2에 대한 역 전사 중합효소 연쇄반응(RT-PCR, 바이러스 유전자를 실시간으로 증폭시키는 방법) 테스트를 통해 광범위하게 이루어진다(Earnest et al., 2022). 코로나-19 진단 시 양성 여부를 결정하는 것은 Cycle Threshold values, 흔히 Ct values로 일컬어지는 측정치이다. Ct values는 유전자 증폭을 몇 차례 거쳤을 때 특정 역치 값(threshold) 이상으로 상승되어 바이러스 감염을 확정할 수 있을 때까지 걸린 증폭 사이클 수를 의미하여 바이러스 부하(viral load)와 반비례한다(Tom MR, Mina MJ, 2020). 코로나-19 진단키트의 Ct values 양성 역치 값은 각 키트별로 표적 유전자(target gene)와 사용 장비의 차이를 고려하여 키트 제조 회사가 결정하므로 고유한 값이 존재한다.

Ct values는 코로나-19 전염성 및 감염력과 연관이 있다는 선행 연구들이 다수 존재하며(Aranha et al., 2021; Marks et al., 2021), 특히 감염된 개인에게서 시간에 따라 체내 바이러스 부하가 줄어듦에 따라 그 값이 커지는 경향을 띤다. Ct values는 이렇듯 바이러스 부하와 반비례한다는 점, 전염성/감염력과의 연관성, 그리고 코로나-19 진단과 동시에 측정 및 기록되는 값으로 즉각적인 활용이 가능하다는 특장점이 있다. 이러한 특장점들로 인해 코로나-19에서 공중 보건 의사 결정을 위한 Ct values의 유용성에 관한 연구들이 전세계적으로 활발히 이루어지고 있다(Waudby et al., 2021; Rodríguez et al., 2021).

대한민국은 코로나-19 이전에도 2009년 신종 플루의 대유행 때 real-time PCR을 대규모로 보급한 이후 계속해서 감염병 진단에 활용해왔다. 덕분에 세계에서 가장 빠른 속도로 진단 키트를 제작하여 코로나-19에 신속하게 대응할 수 있었다(이진수, 2020). 대한진단검사의학회에서 제작한 “코로나바이러스감염증-19 검사실 진단 지침”을 통해 국내에는 코로나19 발생 초기부터 현재까지, 모든 의료기관 및 검사기관에서 시행한 전체 검사자의 Ct values 측정값 정보가 존재한다. 그러나 국내에서는 Ct values가 거의 진단 분야에만

사용되어 왔고, 임상적 특징을 확인하려는 시도는 있었으나(안선욱 등, 2021; Suh et al., 2020) 대상 집단의 수가 적고 그 외에 예측, 전파력 등에 관한 연구는 이루어지지 않았다.

신규 확진 수 예측과 보건 및 감염병 대응전략은 전세계적으로 코로나19 발생 초기부터 현재까지 주간감염재생산지수(time-varying reproduction number, Rt)에 의존해오고 있다(Tizzoni et al., 2022; Hoque et al., 2022). Rt값의 정의는 1명의 확진자가 추가 확진자를 몇 명 만들어냈는지 나타내는 지표로, 그 값이 1을 넘으면 유행이 확산하고 있다는 뜻이다. 그러나 Rt 추정은 일주일 미만으로 줄이기 어렵고(Lin et al., 2022), 그 정의 및 계산식 상 감염경로를 파악해야 정확한 추정이 가능하므로(유명수 등, 2021; 정재웅 등, 2020) 대규모 확진 등으로 확진 환자가 급증하여 역학조사가 제대로 이루어지지 않을 경우 예측 효용성이 저하된다. 예측하지 못한 대규모 확진은 자원 및 인력 부족과 위·중증 및 사망위험을 높일 수 있으므로, Rt의 한계를 보완할 수 있는 새로운 지표로 대한민국 코로나-19 신규 확진 예측 모형을 개선할 필요가 있다.

공중 보건 의사 결정을 위한 Ct values의 유용성은 현재 많은 토론과 논쟁의 주제이며(Hay et al., 2021), Ct values 분포를 활용하여 코로나-19 확진 양상을 예측한 연구들이 국제적으로 활발히 이루어지고 있다. 그 중 다수의 예측 연구에서 Ct value가 Nowcasting 혹은 Rt와 결합하여 확진 양상 예측 모델을 보완하는 형태로 활용될 수 있음을 보였다(Lin et al., 2022; Khalil et al., 2022; Alizon et al., 2022).

본 연구에서는 Ct values 분포가 Rt와 달리 감염경로 확인에 구애받지 않는다는 점에 착안하여, Rt의 한계를 보완할 수 있는 지표임을 확인하고자 하였다. 이를 통해 개선이 필요한 대한민국 코로나-19 확진 양상 예측 모형에 새로운 지표로 Ct values의 활용 가능성을 제시하고자 하였다.

1.2. Systematic review

Ct values를 활용한 코로나-19 모니터링 및 예측 연구 동향을 살펴보기 위하여 체계적 문헌 고찰(Systematic Review)을 수행하였다. 체계적 문헌 고찰은 최신 동향을 반영하기 위해 2020년 2월부터

2022년 9월 30일까지의 총 세 번의 시기(2021년 11월 30일, 2022년 3월 31일, 2022년 9월 31일)에 걸쳐 이루어졌다. 국제 database(DB)는 PubMed, MedLine, Cochrane Library, Embase를 선정하였고 국내 database(DB)는 KISS, RISS, KOREAMED를 선정하였다. 문헌 검색에 사용된 검색식과 문헌 선택 과정은 다음과 같다(Table 1)(Figure 1).

국내 DB에서 69편, 국외 DB에서 11,879편을 합쳐 총 11,948편의 문헌이 검색되었다. 이 중 중복 논문 4,697편을 제거한 후 제목과 초록의 내용을 확인하여 주제와 관련 없는 논문 4,212편, 중재가 부적합한(Ct value를 주로 활용한 것이 아닌) 논문 2,805편을 1차로 제외하였다. 이후 논문 전체 내용을 확인하여 연구의 결과가 부적합한(검사방법 효용(새로운 키트의 유용성 소개 혹은 진단검사의 문제점 분석), 염기서열 분석 등) 논문 130편, 연구 출판범위 부적합(비영어권논문, 학회발표 등) 22편을 2차로 제거하였다. 최종적으로 분석에 포함된 논문은 82편이었다.

최종 선정된 문헌 82편의 내용을 연구 국가, 검체 유형, 변이 바이러스(Variants of Concern, VOCs)포함, 연구대상 시기, 연구대상 수, 표적유전자, 주요 결과로 정리하였다(Appendix table A). 항목에 해당하는 내용을 문헌 및 supplementary files에서 찾을 수 없는 경우는 “NA”로 표기하였다. 체계적 문헌 고찰 결과 총 36개국에서 연구가 이루어져 다양한 국가에서 활발히 Ct values를 통한 코로나-19 모니터링 및 예측 분석을 하고 있다는 것을 알 수 있었다. Ct values를 활용한 연구의 주요 결과는 확진 환자의 임상적 특성에 관한 것이 38편(46.3%)으로 가장 많았으며, 향후 확진 양상 예측과 관련된 경우는 11편 (13.4%), 폐수 및 하수 기반의 모니터링 연구는 14편(17.1%), 유전체 연구 7편(8.5%), 진단 연구와 사회학적 연구는 각각 6편(7.3%)이었다. 또한 58편(70.7%)의 연구가 변이 바이러스 등장 이전 시기에 수행되었고, 61편(74.4%)의 연구가 일부 병원 환자들을 대상으로 한 것으로, 국가 수준에서 변이 바이러스 등장 시기를 포함한 분석한 예측 연구는 2편(2.4%)에 불과하였다.

본 연구와 기존 Ct values 기반 코로나-19 관련 Forecasting 연구들을 비교 및 대조하기 위한 대조표(Table 2)를 작성하여 연구기간, 변이바이러스 포함 여부, 연구대상의 인구집단 대표 여부, 표적유전자,

방법론 및 Ct values를 통한 예측결과에 대해 서술하였다. Ct values를 이용한 지역사회 유행병 양상 예측 및 감시가 가능하며(Hay et al., 2021; Andriamandimby et al., 2022; Walker et al., 2021), 연구 결과는 신규 확진 수 예측 관련이 5편, 코로나-19의 임상적 예후(중증 진행률, 사망률) 예측이 6편이었다. 신규 확진 수 예측 연구에서 사용한 방법론은 정적 추정을 사용한 연구가 2편(Yin et al., 2021; Phillips et al., 2022)과 동적 추정을 사용한 연구가 3편(Lin et al., 2022; Khalil et al., 2022; Alizon et al., 2022)이었으며, 정적 추정을 통한 예측 결과와 동적 추정을 통한 예측 결과에 차이가 존재하였다. 정적 추정 결과는 14일 이상의 시차를 두고 Ct values와 신규 확진 수의 연관성이 추정되었고, 동적 추정 결과는 7일 이하의 시차를 두고 Ct values가 신규 확진 수를 예측하였다.

체계적 문헌 고찰 결과, 변이 바이러스 등장 시기를 포함한 인구집단 수준의 Ct values 분포를 활용한 코로나-19 예측(Forecasting) 연구가 미비하며, 사용한 추정 기법에 따라 예측 결과에 차이가 있음을 알 수 있었다. 이에 본 연구에서 인구집단 수준의 국내 Ct values 데이터로 정적 추정과 동적 추정 기법을 모두 활용한 신규 확진 예측 연구를 수행하여 이러한 미비점을 보완하고자 하였다.

Table 1. Search terms used in national/international database

International DB Keyword (Search date: until Sep 30th, 2022)	
#1	TI, AB : "surveillance" OR "monitor*" OR "model*" OR "trajectory" OR “predict"† OR “forecast*”§
#2	TI, AB : “PCR”¶ OR “antigen”¶ OR "cycle threshold" OR “Ct values" OR “RdRp”‡ OR "viral load“ OR “reproduction number”‡ OR “biomarker”‡ OR “Cq values"§
#3	TI, AB : "Coronavirus" OR "COVID-19" OR “covid“ OR "SARS-CoV-2"
International DB > PUBMED, MedLine¶, Cochrane Library§, Embase	

National DB Keyword (Search date: until Sep 30th, 2022)	
#1	초록: surveillance | monitor | monitoring | model | modeling | trajectory | predict | forecast§ | 감시 | 모니터링 | 모델 | 추적 | 예측†
#2	초록: PCR¶ | antigen¶ | cycle threshold | Ct values | RdRp‡ | viral load | reproduction number‡ | biomarker‡ | 항원¶ | 바이럴 로드 | 바이러스 배출량 | 감염재생산‡ | 바이오마커‡ | Ct값
#3	초록: Coronavirus | COVID-19 | covid | 코로나 | 바이러스 | SARS-CoV-2
National DB > KISS, RISS, KOREAMED	

†: 2021/11/30 이후 검색에 추가됨
‡: 2021/11/30 이후 검색에 포함되지 않음
§: 2022/3/31 이후 검색에 추가됨
¶: 2022/3/31 이후 검색에 포함되지 않음

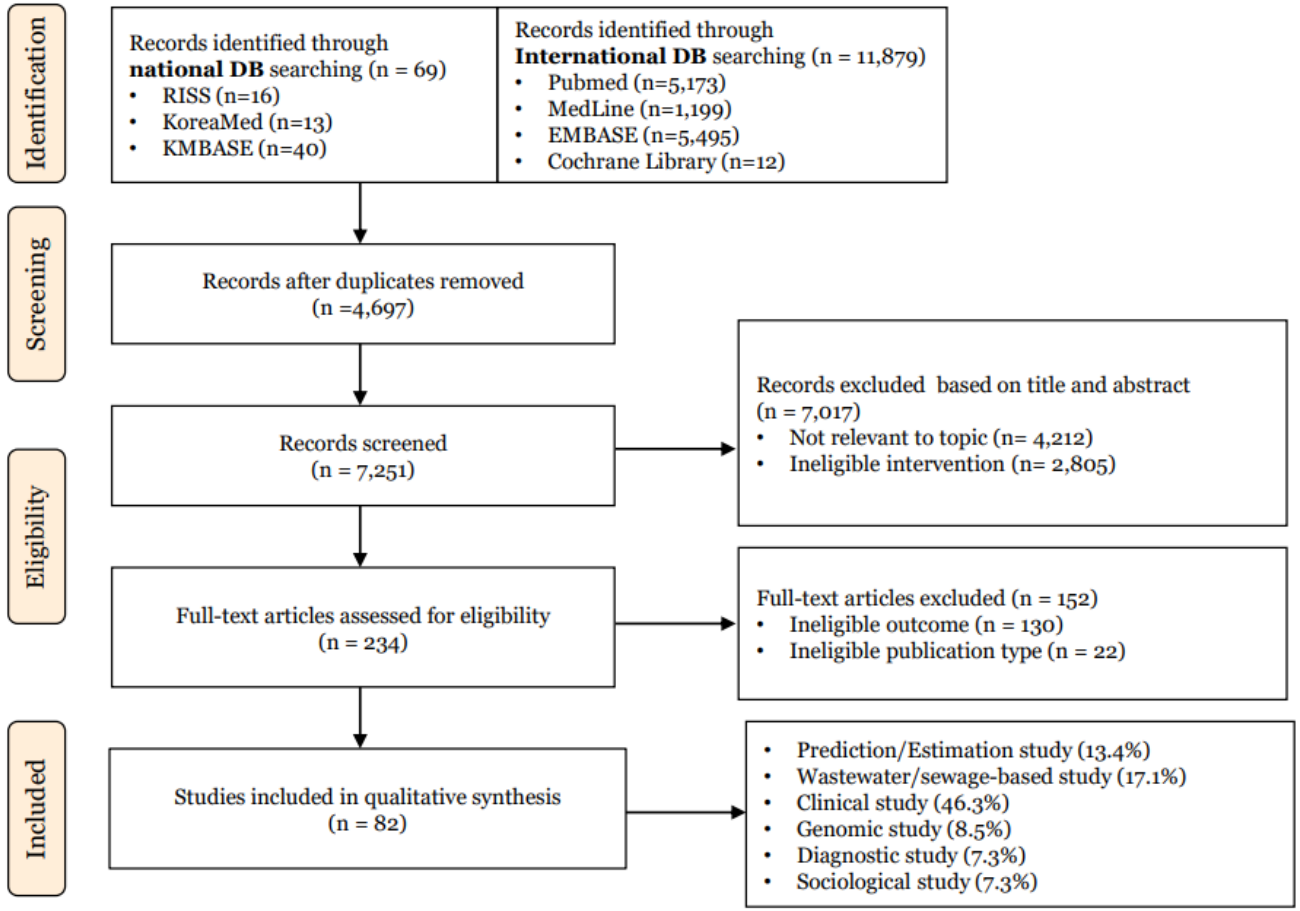


Figure 1. Flow diagram of literature and selection criteria

Table 2. Comparison with previous forecast studies

Authors	변이바이러스 환자 포함	연구기간이 1년 이상	인구집단 수준의 study sample	CT value target gene	Ct value 예측	방법론
Yin, N. 2021	✓	✓	✓	RdRp, N	Ct value median 값은 17일 lag로 일별 검사 양성과 연관이 있음	스피어만의 순위상관계수를 통한 상관성 분석
Skarzynski, M. 2021	✓	✓	X	N1	Ct value의 중위값이 16.6 이하일 경우, 21.4 이상인 경우보다 임상적 예후가 악화될 확률이 현저히 높았음	피어슨 상관계수를 통한 상관성 분석, genome-based predictive algorithm
Ingberg, E. 2022	X	X	X	N, E, RdRp	COVID-19 환자에서 흉부 CT를 통해 확인한 폐 침엽 정도와 Ct value의 연관성을 확인하고, 두 가지 지표를 결합하여 질병의 중증 진행률 예측	다중 로지스틱 회귀분석, 다중 선형회귀모형
Phillips, M. C. 2022	✓	X	X	E, N2	Weekly-average Ct value는 향후 2주 동안 확진자 수와 가장 밀접하게 연관됨	단순 선형회귀모형
Lin, Y. 2022	X	X	✓	E	Now-casting	SEIR 모형, EpiNow2
Khalil, A. 2022	X	✓	✓	NA	Now-casting	스피어만의 순위상관계수를 통한 상관성 분석, RNN(Recurrent Neural Network) models
Hay, J. A. 2021	✓	X	NA	N1, N2, ORF1ab	Ct값의 중위값과 왜도를 통해 epidemic 양상 예측할 수 있음	various SEIR models, mathematical model for population-level viral load distributions
Barry, M. 2022	X	X	X	E	RT-PCR Ct 값은 IMV(invasive mechanical ventilation)와 사망률의 예측 변수로 사용할 수 있음	단변량 로지스틱 회귀모형, 다변량 로지스틱 회귀모형
Andriamandimby, S. F. 2021	X	X	✓	E, N, ORF1a/b, or S	Ct value가 낮을 때 유행병이 최고의 증가율을 보였으며, Ct value를 이용해 개인이 감염된 이후 경과한 시간에 대한 예측이 가능하다는 것을 확인함	GAM(Generalized additive modeling), SEIR model
Alizon, S. 2022	X	X	✓	NA	Ct value median 값은 6-7일 lag로 Rt값과 연관이 있음	선형회귀모형, ARIMA model
Walker, A. S. 2021	✓	X	✓	N, S, ORF1ab	Ct value를 이용한 지역사회 유행병 감시체계의 가능성을 확인.	스피어만의 순위상관계수를 통한 상관성 분석, median(quantile) regression
The proposed model (current study)	✓	✓	✓	RdRp	Ct value median 값은 특정 lag 일수로 일별 신규 확진 수를 예측할 수 있음	단순 선형회귀, FDL model(유한시차모형)

1.3. Study objective

연구 목표는 다음과 같다. 먼저 코로나-19 유행 시기별 Ct value 시계열 분포를 통해 Rt와 Ct value를 비교하고, Ct value가 실제 신규 확진을 예측할 수 있을지 시각적으로 확인하였다. 그 다음, 코로나-19 유행 시기별 Ct value를 활용한 코로나 확진 양상에 대한 시계열 분석 기반 예측을 시차가 반영된 정적, 동적 추정 기법을 모두 활용하여 몇일의 시차를 두고 가장 예측력이 높은 지 확인하였다.

2. METHODS

2.1. Study design

이 관찰 단면 연구는 역학에서 관찰 연구의 보고 강화(STROBE) 진술에 따라 수행되고 문서화되었다.

Study site

국가기관인 질병관리청(협조 당시 질병관리본부)과 대한진단검사의학회가 공동으로 주관하여 2020년 2월 7일 ~ 2021년 12월 31일까지 약 2년간 대한민국의 총 182개 의료기관 및 검사전문기관에서 수행된 전 국민 대상 코로나-19 진단 결과(코로나-19 감염 의사환자, 감염 확진자 접촉자, 또는 원인불명 폐렴으로 진단받고 코로나-19 real-time RT PCR을 시행한 환자의 의무기록)를 대상으로 하였다.

Specimen collection and testing

11개 회사의 진단키트 제품(SD 바이오센서, SML 제니트리, 랩지노믹스, 로슈, 바이오니아, 바이오세움, 바이오코아, 씨젠, 오상헬스케어, 캔서롭, 코젠바이오텍)을 이용한 진단검사가 시행되었다.

진단검사는 코로나-19 유전자(Real-time RT-PCR) 검사를 의미한다. 국내 RT-PCR 프로토콜은 질병관리청과 WHO의 지침을 기준으로 E 유전자를 선별 검사(screening test), RdRp 유전자를 확인 검사(confirmatory test)로 사용하는 2단계 해석 알고리즘을 사용하므로, 본 연구에서는 Ct value RdRp를 대표 Ct value로 분석에 사용하였다.

확진환자(Confirmed case)는 임상양상에 관계없이 진단을 위한 검사기준에 따라 감염병 병원체 감염이 확인된 자를 의미하며, 본 연구에서는 확진환자 중 신규 확진 수 예측이라는 목적에 맞게 기확진 환자를 제외하고 신규 확진 환자(신환)만 확진 환자로 정의하였다.

코로나-19 유행 시기 구분

질병관리청에서 발간한 “주간 건강과 질병·제15권 제7호(2022. 2. 17.) 국내 COVID-19 확진자 2년 발생 보고서”를 기준으로 국내 코로나-19 유행 시기를 구분하였다.

제1기는 2020년 1월 20일부터 8월 11일까지이며, 제2기는 2020년 8월 12일부터 11월 12일, 제3기는 2020년 11월 13일부터 2021년 7월 6일까지, 제4기는 2021년 7월 7일부터 2022년 1월 19일까지에 해당한다.

2.2. Data source

대한진단검사의학회 코로나-19 양성결과 자료

2020년 2월 7일 ~ 2021년 12월 31일까지 대한진단검사의학회 신종 코로나바이러스 대책위원회에서 발간한 “코로나바이러스감염증-19 검사실 진단 지침"에 따라 코로나-19 진단 검사를 수행하고 입력된 자료를 사용하였다. 학위논문 연구목적에 맞게 대한진단검사의학회로부터 자료제공 허가와 1차 자료 연구책임자로부터 제공 심의를 통해 자료를 제공받았다. 검사기관, 검사일, 기관시도, 연령, 성별, 검사키트 정보, 검체 종류, 표적유전자 Ct value, 환자구분에 관한 변수를 포함하며, 사용할 자료는 개인식별정보가 제거된 상태였다.

서울대학교 아시아연구소 COVID-19 dataset

서울대학교 아시아연구소(SNUAC)의 아시아지역 정보센터(ARIC) 홈페이지에서 이용 가능한 공개된 자료인 국내 코로나-19 자료(‘USABLE KOREA DATA: COVID_KOREA’)로, 2020년 2월 7일 ~ 2021년 12월 31일까지 일별 확진 현황 자료를 추출하여 사용하였다.

Our World in Data COVID-19 dataset

Our World in Data 홈페이지에서 이용 가능한 공개 자료인 ‘국제

코로나-19' 자료로, 2020년 2월 21일 ~ 2021년 12월 31일까지 대한민국 일간 감염 재생산지수(Rt) 값을 추출하여 사용하였다.

2.3. Sample size

연구 대상 검체 선택 과정은 다음과 같다(Figure 2). 총 884,853건의 코로나-19 양성 진단 결과 자료 중 신환(첫 검체로 초진)이 아닌 경우 314,886건, 양성 판독 수치 부적합(키트 별 양성으로 판독하는 수치에 벗어난 Ct value 측정치 기록사례)으로 위양성 의심 사례 23,535건이 제외되었다. 이후 본 연구에서 사용하는 Ct value RdRp 값이 누락된 경우 34,648건과 연령을 1-100세로 한정하여 21,926건을 제거하였다. 마지막으로, Ct value는 진단 키트별로 고유의 양성 판독 수치가 존재하므로 사용된 진단 키트를 알 수 없는 경우 735건을 제거하여 최종적으로 489,133건이 분석에 포함되었다.

성별, 핵산추출방식, 유전자증폭장비, 검체채취부위, RdRp 이외의 Ct value 변수에도 missing이 있었으나, 이러한 결측을 모두 삭제하면 오히려 유효한 sample size가 과하게 줄어들어 "Ct value의 분포를 활용한다"는 연구목표를 저해할 가능성이 있었다. 그러므로 이러한 결측을 남겨 두되, 각 변수에 따른 Ct value 분포를 확인하는 분석 시에는 각 변수에 존재하는 결측을 제외하고 분석하였다.

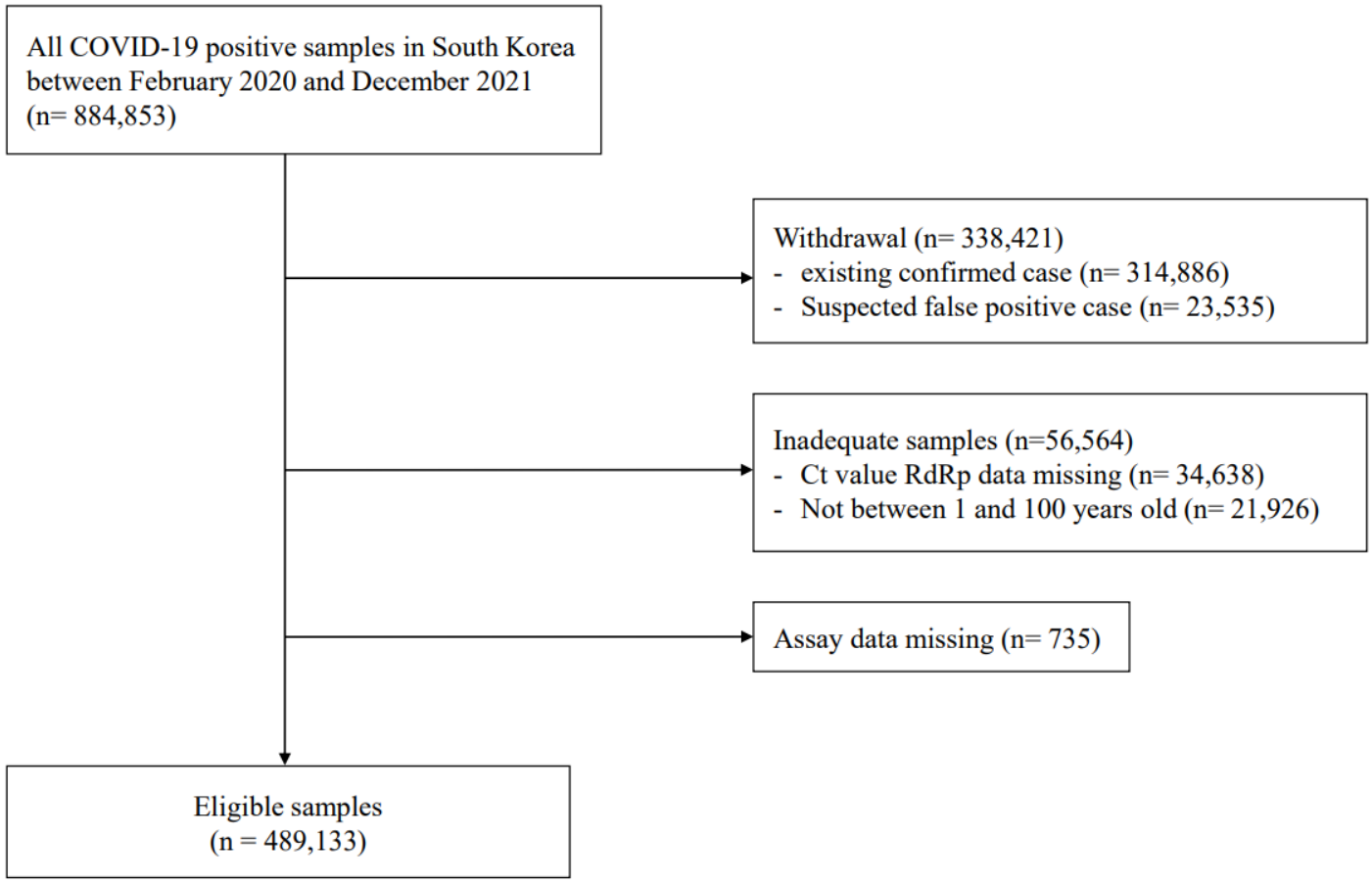


Figure 2. Flow diagram of sample selection

2.4. Statistical analysis

Descriptive statistics for diagnostic test results

먼저 분석에 포함된 연구 대상 검체의 진단검사 결과를 전반적으로 파악하기 위해 진단검사 결과의 각 변수에 대한 기술통계 표 및 그래프를 제시하였다. 대상자의 인구학적 정보, 계측 정보 및 검사실 검사 결과를 연속형 변수는 평균 ± 표준편차(SD), 4분위수(25%, 50%, 75%로 나타내고 범주형 변수는 빈도 및 퍼센트로 나타내었다. 여러 Ct values의 분포를 한 눈에 파악하기 위해 violin plot을 사용하였다.

Distribution patterns of Ct values in Korea

신규 확진 예측에 Ct value의 활용을 재고하기 위해 그 분포를 다른 수치들과 파악하여 분석 이전의 시각적 확인을 할 필요가 있었다. 두 가지의 그래프를 통해 코로나-19 유행 시기별 인구집단 수준의 Ct value 분포 양상을 각각 Rt(감염재생산지수)와 신규 확진 수와 함께 시계열 그래프로 나타내었다.

첫 번째 그래프에서는 유행 시기별 Ct value, Rt 분포를 비교하였다. Ct value는 진단검사 개인 결과를 주별(weekly)로 통합한(aggregated) median 값을, Rt는 Our World in Data 공개자료에서 얻은 일별 Rt 값을 주별(weekly)로 통합한 값을 사용하였다. 이를 통해 Ct value가 Rt와 비교해 예측할 만한 가치가 있는지, 있다면 어느 시기에 더 예측 효용이 좋을지 시각적으로 확인하였다.

그 다음 두 번째 그래프로 유행 시기별 Ct value, 신규 확진 수가 함께 나타난 분포에 방역정책 강화/완화 시기를 표시하였다. Ct value는 진단검사 개인 결과를 일별(daily)로 통합한 median 값을, 신규 확진 수는 서울대학교 아시아연구소(SNUAC) 공개자료에서 얻은 일별 신규 확진 수 자료를 사용하였다. 이 그래프로 Ct value가 신규 확진을 예측할 수 있는 지표인지, 방역정책 양상이 변화함에 따라 영향을 받는지 확인하였다.

Forecasting the number of newly confirmed cases based on the Ct values

앞서 두 가지 분포 그래프를 통해 확인한 Rt의 한계와 Ct의 신규 확진 예측지표로의 효용을 확인한 후, Ct value의 실제 예측력을 확인하기 위한 통계적 분석을 진행하였다. Ct value가 몇일의 시차를 두고 가장 예측력이 높은 지 확인하기 위해 신규 확진 수를 예측할 때, time-lag(시차)를 반영한 두 가지 분석 기법을 사용하였다. 하나는 단순 선형회귀모형, 다른 하나는 Distributed Lag Model(시차분포모형) [DLM]이다.

먼저 정적(static) 추정 기법인 단순 선형회귀모형으로 독립변수인 Ct value에 time lag를 주어 종속변수인 신규 확진 수가 어느 시차에 가장 유의하게 예측되는지 확인하였다. 이후 시차에 따른 Adjusted R-sqaure 값과 회귀 계수의 변화를 확인하여 시간 변화에 따른 효과가 있는지 파악하였다. 모형에서 시차(time lag)는 선행 연구를 통해 밝혀진 코로나-19 바이러스의 최대 잠복기인 14일(Li et al., 2020)까지 반영하였다.

그 다음, 시계열 자료 특성에 맞는 동적(dynamic) 추정 기법인 DLM을 사용하여 시간적 변동에 따른 효과를 제거하고, 며칠까지의 시차에 가장 유의하게 신규 확진 수가 예측되는지 확인하였다. DLM은 시계열 회귀 모형 중 유한 개의 독립변수(예측변수, 설명변수)의 현재값과 과거값(lagged variables)을 이용하여 종속변수(예상변수, 반응변수) y를 예측하는 동적 추정 기법이다(Hamilton, 1994). 시계열 자료에 대한 회귀 예측은 주로 ARIMAX 모형(Autoregressive Integrated Moving Average Exogenous Model) 으로 수행되는데, 해당 모형은 독립변수로 종속변수의 시차변수가 포함되고 외생 변수로 X가 존재하는 모형이다(나종화, 2020). 본 연구의 목적 상 Ct values의 현재값 및 과거값이 독립변수에 포함되고 종속변수는 신규 확진 수로 규정되므로 연구목적에 부합하는 시계열 회귀 모형인 DLM을 사용하였으며, 수식은 다음과 같다.

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_n x_{t-n} + \varepsilon_t, \quad t = n+1, \ldots, T$$

시차 T를 유한하게 규정할 경우 DLM은 유한시차모형(Finite

Distributed Lag model; FDL model)이 된다. 본 연구에서는 선형회귀분석과의 비교를 위해 동일한 시차인 14일로 모형을 규정하였으므로, DLM 분석 이전에 시차를 결정하기 위한 교차상관분석(Cross-Correlation Function analysis, CCF)은 수행하지 않았다. 수식에서 추정계수 β_0은 충격승수(impact propensity, impact multiplier)로, 독립변수의 현재 값이 종속변수의 현재 값에 미치는 효과를 의미한다. 추정계수 β_1과 β_2, … β_n은 지연승수로, 독립변수의 과거 값이 종속변수의 현재 값에 미치는 효과(독립변수의 현재 값이 종속변수의 미래 값에 미치는 효과)를 의미한다. 따라서 독립변수가 1 단위 변화할 때 종속변수의 모든 미래값에 영향을 미치는 장기효과(Long Run Propensity(LPR))는 해당 시차까지의 모든 승수를 더한것인 ($\beta_0+\beta_1+\beta_2$ +…+ β_n) 로, 총승수(total propensity)로도 불린다(Jenkins, & Box, 1976). 시간적 변동에 따른 효과를 제거하여 정상성(Stationarity)을 확보하기 위해 사전 검정으로 Augmented Dickey-Fuller test(ADF) 기법을 사용한 단위근 검정(Unit root test)을 한 후, 비정상 시계열은 로그 변환 및 차분(Differencing)을 통해 정상 시계열로 변환하여 분석을 진행하였다.

두 모형 모두 Ct value는 진단검사 개인 결과를 일별(daily)로 통합한 median 값을, 신규 확진 수는 서울대학교 아시아연구소(SNUAC) 공개자료에서 얻은 일별 신규 확진 수 자료를 사용하였다. 또한 두 모형 모두 전체 시기, 유행 시기(제1기, 제2기, 제3기, 제4기)에서 각각 분석을 진행하여 어느 시기에 가장 예측이 잘 되었는지 확인하였다. DLM(FDL model)을 통한 예측 분석을 수행한 이후, 예측 모형이 잘 적합 되었는지 판단하기 위한 사후 검정으로 Breusch-Godfrey test 기법을 사용한 자기상관 검정(Autocorrelation test)을 수행하였다.

p 값이 0.05 미만인 경우를 통계학적으로 유의하다고 판단하였으며, 모든 그래프 및 통계분석은 STATA(버전 17)로 수행하였다.

Ethics Statement

This study was approved by Seoul National University institutional review board (IRB No. E2206/004-001).

3. RESULTS

3.1. Descriptive statistics for diagnostic test results

분석에 포함된 연구 대상 검체의 진단검사 결과를 전반적으로 파악하기 위해 진단검사 결과의 각 변수에 대한 기술통계 표(Table 3)를 구성하였다.

검사기관, 연령 및 성별에 대한 결과로는 총 489,133건의 양성 건수 중 대부분인 446,653건(91.32%)이 검사전문기관에서, 42,480건(8.08%)이 의료기관에서 이루어졌다. 검체 채취 시/도로 구분한 결과, 서울특별시가 202,328건(41.36%)으로 가장 높았고, 다음 순위로는 경기도가 147,503건(30.16%), 인천광역시가 56,017건(11.45%), 부산광역시가 37,511건(7.67%)으로 그 뒤를 이었고, 가장 적은 곳은 전라남도로 28건이 전부였다. 그러나 이는 검사기관 및 의료기관의 소재지에 의한 것으로, 확진 환자들의 실제 거주지/발생지와는 차이가 존재한다. 연령 평균은 42.05세, 4분위수는 각각 26세(25%), 42세(50%), 59세(75%) 였다. 성별은 245,147건(50.12%)이 남성, 223,741건(45.74%)이 여성으로 남성이 약간 더 많았다.

키트에 관련된 결과는 다음과 같다. 사용된 진단키트는 ㈜씨젠이 293,004건(59.90%)으로 절반을 넘게 차지하였고, 다음 순위로는 ㈜바이오세움이 97,250건(19.88%), SD바이오센서㈜가 75,664건(15.47%), ㈜코젠바이오텍이 22,049건(4.51%)으로 그 뒤를 이었다. 유전자 증폭장비는 CFX96 Real-time PCR이 거의 모든(99.16%) 검사에서 사용되었으며, 검체의 종류는 인후도말(NPS/OPS)이 246,722건(50.44%)으로 가장 많이 사용되었고, 나머지는 비인두도말(NPS)이 224,198 건(45.84%), 객담이 13,949(2.85%)건으로 대부분 사용되었다.

Ct value에 관련된 결과는 Ct value E, RdRp, N의 분포를 한 눈에 파악할 수 있는 violin plot을 통해 나타냈다(Figure 3). Violin plot은 커널 밀도 곡선(Kernel Density Curve)와 box plot을 합친 형태로 단일 변량, 연속형 데이터의 분포를 카테고리 값에 따른 각 분포의 실제

데이터 또는 전체 형상을 보여준다는 장점이 있다(Hintze, J. L., & Nelson, R. D., 1998). Violin plot을 통해 세 가지 Ct value가 모두 비슷한 분포 형태를 지니며, 값이 10-20 사이에 주로 많이 분포하는 특징 또한 관찰되었다.

각 진단검사 결과 변수 별 Ct value의 시계열 분포 그래프는 Appendix B에서 확인할 수 있다.

Table 3. Descriptive statistics for diagnostic test results of samples

	N (%)	Mean (±SD)	Median ($25^{th}-75^{th}$ percentiles)
검사기관 구분	489,133 (100.00)		
검사전문기관	446,653 (91.32)		
의료기관	42,480 (8.68)		
검사기관 시도	489,133 (100.00)		
강원	517 (0.11)		
경기	147,503 (30.16)		
경남	686 (0.14)		
경북	277 (0.06)		
광주	12,524 (2.56)		
대구	2,151 (0.44)		
대전	1,910 (0.39)		
부산	37,511 (7.67)		
서울	202,328 (41.36)		
울산	158 (0.03)		
인천	56,017 (11.45)		
전남	28 (0.01)		
전북	4,037 (0.83)		
제주	167 (0.03)		
충남	1,255 (0.26)		
충북	22,064 (4.51)		
검사일	489,133	22511.58 (± 146.87) [21.08.19]	22559 (22473 – 22619) [21.10.06] ([21.07.12] – [21.12.05])
연령	489,133	42.05 (± 21.15)	42 (26 – 59)
성별	489,133 (100.00)		
남	245,147 (50.12)		
여	223,741 (45.74)		
Missing	20,245 (4.14)		
진단키트 제조사	489,133 (100.00)		
SD 바이오센서	75,664 (15.47)		
SML제니트리	113 (0.02)		
랩지노믹스	95 (0.02)		
로슈	265 (0.05)		
바이오니아	66 (0.01)		
바이오세움	97,250 (19.88)		
바이오코아	13 (<0.01)		
씨젠	293,004 (59.90)		
오상헬스케어	254 (0.05)		
캔서롭	360 (0.07)		
코젠바이오텍	22,049 (4.51)		
유전자 증폭장비	489,133 (100.00)		
Applied Biosystems 7500 Fast Real-time PCR	747 (0.15)		
Applied Biosystems 7500 Real-time PCR	2,557 (0.52)		
CFX96 Real-time PCR	485,046 (99.16)		
cobas 6800/8800	263 (0.05)		

	N (%)	Mean (±SD)	Median ($25^{th}-75^{th}$ percentiles)
Missing	520 (0.12)		
검체종류	489,133 (100.00)		
BAL(기관지폐포세척액)	1 (<0.01)		
NPA(비인두흡인액)	8 (<0.01)		
NPS(비인두도말)	224,198 (45.84)		
NPS/OPS	246,722 (50.44)		
OPS(구인두도말)	77 (0.02)		
TA(기관흡인액)	6 (<0.01)		
객담	13,949 (2.85)		
타액	7 (<0.01)		
Missing	4,165 (0.85)		
Ct value E	488,912	21.18 (± 6.52)	20.06 (15.91 - 26.29)
Ct value RdRp	489,133	21.77 (± 6.61)	20.58 (16.42 - 26.97)
Ct value N	293,545	21.60 (± 6.75)	20.52 (16.17 - 26.92)
Ct value ORF1ab	0	no observations	no observations
Ct value S	0	no observations	no observations

***Data are mean (±SD), median ($25^{th}-75^{th}$ percentiles) or n (%).**
Assays that target RdRp gene don't target ORF1ab or S gene.

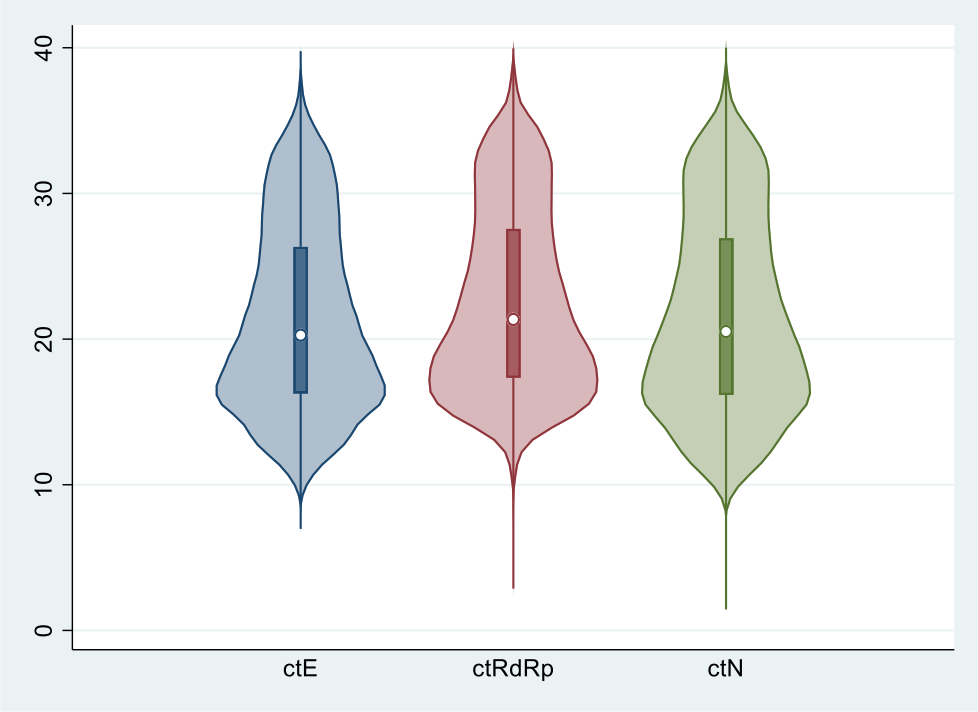


Figure 3. Violin plot of Ct value E, RdRp, and N

3.2. Distribution patterns of Ct values in Korea

Time series distribution graph of Ct values RdRp and Reproduction rate (Figure 4)

Ct value가 Rt와 비교해 예측할 만한 가치가 있는지 확인하기 위해, 코로나-19 유행 시기별 Ct value의 분포를 Rt(감염재생산지수)와 함께 시계열 그래프로 나타냈다. Ct value의 축 척도는 Rt와 직관적으로 비교하기 위해 역-척도를 사용하여, 값이 낮을수록(viral load가 높을수록) 그래프가 상승하는 형태를 띤다.

전체 시기에서 Rt와 Ct value의 그래프 형태는 비슷한 모양을 띠므로, Ct value가 새로운 신규 예측 지표로 활용될 가능성을 나타낸다.

제1기는 2020년 1월 해외유입 확진 발생을 시작으로 대구, 경북 지역 신천지 관련 및 의료기관, 종교시설 등에서 확진 사례가 증가한 시기를 의미한다. 이 시기의 확진 양상은 주요 감염경로가 신천지 관련(35.7%) 이었으며 그 외에 지역집단발생(26.9%), 해외유입(17.6%), 지역사회접촉 1,406명(9.6%) 순으로(양성찬 등, 2022) 감염경로 파악이 잘 되었던 시기이다. 이 시기의 Rt와 Ct value를 비교해 보면, 1기 중반 이전에는 Rt의 경우 대체로 역치 값인 1보다 낮은 상태를 유지하며 상승-하락을 반복하는 형태를 보인다. Ct value 또한 viral load가 대체로 낮은 값(Ct value>25)(Aranha et al., 2021)을 유지하면서 상승-하락을 반복하는 형태를 보인다. 1기 중반 이후에는 Ct value와 Rt 모두 1기 중반 이전에 비해 절대적으로 상승된 형태(Ct value의 값은 하락할 경우 그래프 형태가 상승)를 띤다. 따라서 Ct value와 Rt가 매우 유사한 형태를 띠는 시기로 판단할 수 있다.

제2기는 수도권 종교시설 및 도심 집회 등으로 소규모 및 중간규모 집단이 다수 발생한 시기를 의미한다. 이 시기의 확진 양상은 주요 감염경로가 지역집단발생(55.9%), 지역사회접촉(18.3%), 감염경로 조사중(14.5%) 순으로(양성찬 등, 2022) 1기에 비해 감염경로가 파악되지 않은 경우가 14.5%로 증가된 시기이다. 이 시기의 Rt와 Ct value를 비교해 보면, Ct value가 viral load가 대체로 높은 값(Ct value≤25)을 유지하는 데 비해 Rt는 역치 값인 1보다 낮은 시기가 다소 길다. 따라서 이 시기에서는 Ct value와 Rt 형태 간에 약간의

차이가 존재한다.

제3기는 수도권을 중심으로 교정시설, 의료기관, 종교시설 등 집단이 다수 발생한 시기를 의미한다. 이 시기의 확진 양상은 주요 감염경로가 지역집단발생(38.2%), 지역사회접촉(34.5%), 감염경로 조사중(22.5%) 순으로(양성찬 등, 2022) 2기에 비해 감염경로가 파악되지 않은 경우가 22.5%로 더욱 증가된 시기이다. 이 시기의 Rt와 Ct value를 비교해 보면, Ct value는 3기 초반을 제외하면 viral load가 높은 값(Ct value≤25)을 유지하면서도 평균적으로 1기, 2기에 비해 viral load가 더 높은 값(Ct value≤23)을 띤다. 그에 반해 Rt는 앞선 1기, 2기에 비해서 역치 값인 1에서 크게 변동이 없다. 따라서 이 시기에서는 Ct value와 Rt 형태 간에 차이가 뚜렷하게 존재한다.

제4기는 수도권 및 대도시 중심의 확진자 발생이 지속되었으며, 예방접종력 효과 감소, 델타, 오미크론 등 변이바이러스의 출현 및 우점화로 확진자 수가 증가한 시기를 의미한다. 이 시기의 확진 양상은 주요 감염경로가 지역사회접촉(46.4%), 감염경로 조사중(32.8%) 지역집단발생(18.6%) 순으로(양성찬 등, 2022) 감염경로가 파악되지 않은 경우가 32.8%로 더욱 증가되었을 뿐만 아니라, 주요 감염경로의 두 번째를 차지하여 3명 중 1명 꼴로 감염경로 파악이 되지 않은 시기이다. 이 시기의 Rt와 Ct value를 비교해 보면, Ct value는 viral load가 제1기부터 지속적으로 증가하여 그 어느 시기때보다도 높은 값(Ct value≤20)을 유지하는 형태를 띤다. 반면 Rt는 3기에 비해서는 변동이 더 크고 역치 값인 1보다 대체로 높은 값을 유지하지만, 이는 역치 값인 1에서 크게 변동이 없다.

제1기부터 4기까지 일 평균 확진 수는 꾸준히 증가하였고(양성찬 등, 2022), 감염 경로가 파악되지 않는 경우도 꾸준히 증가하였다. 현재 신규 확진의 예측 지표로 사용되는 Rt는. 꾸준히 증가하는 확진 수를 반영하기에 한계를 보이며, 감염 경로가 파악되지 않는 경우가 증가함에 따라 이 한계가 더욱 명확하게 확인된다. 특히 제3기, 4기에서 4차 대유행과 신규 확진 수가 급증하던 시기가 분명 존재하였는데도 Rt 값의 변동이 심하지 않아 이는 확진 양상을 제대로 반영하지 못하고 있다고 볼 수 있다. 이는 서론에서 Rt의 정의 및 계산식 상 감염경로가 파악되지 않으면 정확한 추정이 불가능하다는 것과 일맥상통한다. 반면 Ct value의 경우, 전체적인 상승/하락의 형태가 Rt와 매우 유사하면서도

감염경로 파악과 관계없이 제1기부터 4기까지 viral load가 꾸준히 증가하여 일 평균 확진 수의 양상을 반영한다고 할 수 있다.

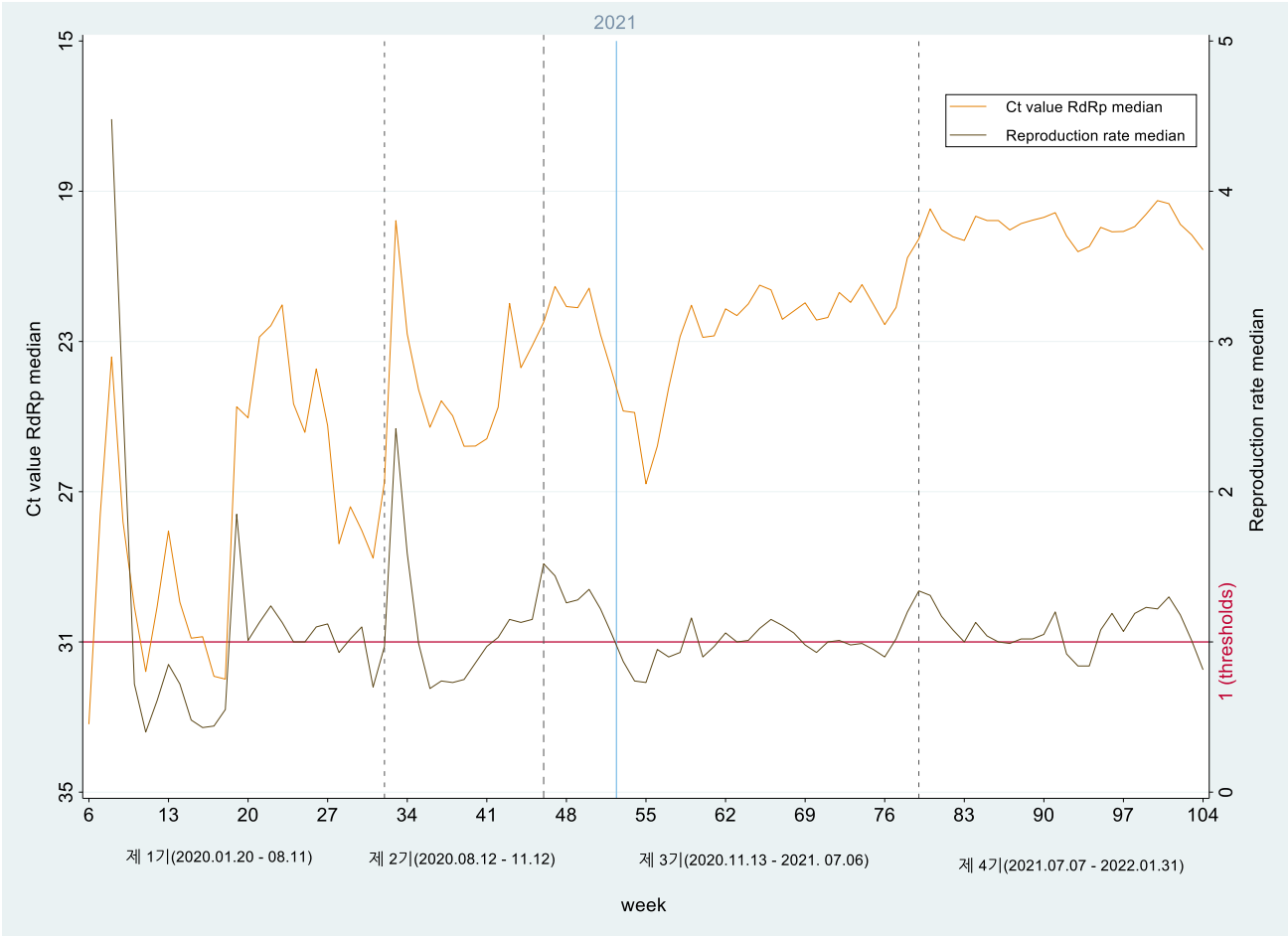


Figure 4. Time series distribution graph of Ct value RdRp and Reproduction rate (by week)

Time series distribution graph of Ct value RdRp and number of newly confirmed cases (Figure 5)

앞서 Ct value가 Rt와 비교하여 확진 양상을 반영하는 지표가 될 수 있다는 것을 확인하여, 나아가 실제 확진 수를 예측할 수 있을지, 방역정책 양상이 변화함에 따라 영향을 받는지 확인할 필요가 있다. 코로나-19 유행 시기 별 Ct value의 분포를 신규 확진 수 뿐만 아니라 방역정책(사회적 거리두기 정책 변화 및 예방접종 정책 도입 등) 강화/완화 시기를 함께 표시한 그래프를 생성하였다. 붉은 영역은 방역정책이 강화되어 있는 시기, 푸른 영역은 방역 정책이 완화되어 있는 시기를 나타낸다.

제1기에 해당하는 붉은 영역은 2020년 3월 22일부터 시행된 강화된 거리두기/5월 29일부터 (수도권)강화된 방역조치 시행에 해당하며, 푸른 영역은 4월 20일부터 일부 조치 완화, 5월 7일부터 시행된 생활 속 거리두기의 주요 방역 정책들이 반영되어 있다.

제2기에 해당하는 붉은 영역은, 2020년 8월 28일부터 시행된 거리두기 2.5단계 강화, 푸른 영역은 10월 11일부터 시행된 거리두기 1단계로 완화의 주요 방역 정책들이 반영되어 있다.

제3기에 해당하는 붉은 영역은, 2020년 12월 8일부터 시행된 거리두기 2.5단계로 격상 및 특별방역조치 실시, 2021년 1월 4일부터 시행된 전국 5인이상집합금지 조치, 2월 26일부터 예방접종 시작에 해당하며, 푸른 영역은 2기에서 이어져온 2020년 10월 11일부터 시행된 거리두기 1단계로 완화/2021년 4월 18일에 발생한 국내 첫 델타변이 감염자부터의 기간에 해당한다.

제4기에 해당하는 붉은 영역은 2021년 7월 12일부터 시행된 수도권 거리두기 4단계로 격상, 10월 12일부터 3차 예방접종 시작, 10월 23일 전국민 예방접종 70% 달성에 해당하며, 푸른 영역은 11월 1일부터 시행된 단계적 일상회복 기간에 해당한다.

모든 시기에서 신규 확진 수와 Ct value 값을 비교하면 큰 폭의 신규 확진 수 상승 이전에 Ct value가 먼저 하락하는 경향을 보였다. 또한 대체로 거리두기 강화 등 방역정책을 강화하면, 신규 확진 수에 영향이 반영되기 이전에 Ct value 값이 먼저 상승하는 현상이 관찰되었다. 반대로 거리두기 단계를 격상하거나 변이 바이러스가

등장하면 Ct value 값이 증가하고 뒤이어 일별 신규 확진 수가 감소하는 형태가 관찰되었다.

이를 통해 Ct value가 확진 수 변동 이전에 값이 변한다는 것과, 방역정책 변화가 확진 수에 영향을 미치기 이전에 Ct value가 먼저 영향을 받아 값이 변동한다는 것을 확인하였다.

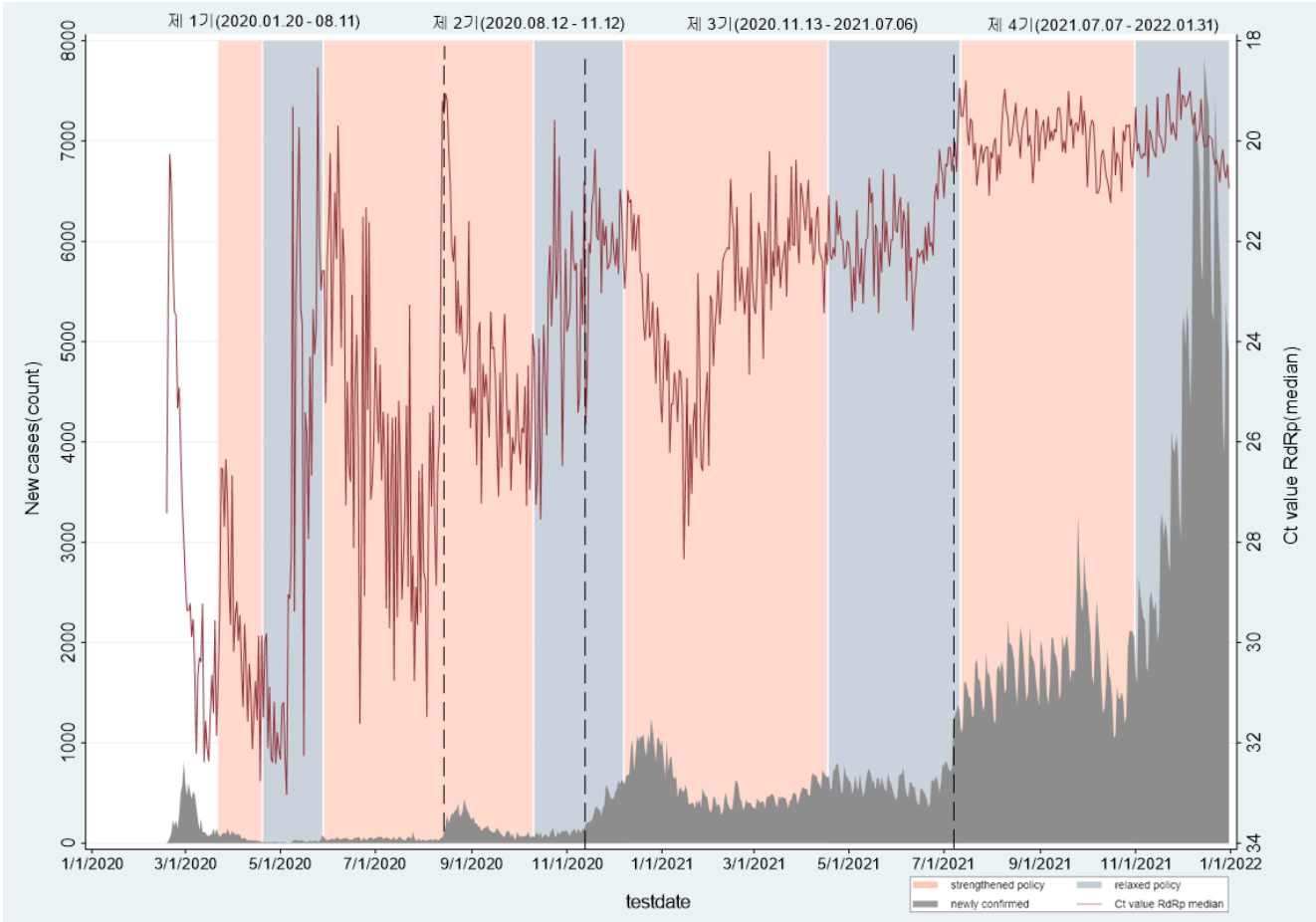


Figure 5. Time series distribution graph of Ct value RdRp and number of newly confirmed cases (by date)

3.3. Forecasting the number of newly confirmed cases based on the Ct values

3.3.1. Static model: Simple Linear regression model

코로나-19 유행 시기별 Ct value와 Rt, 신규 확진자 발생의 분포를 비교한 결과 Ct value가 Rt에 비해 대유행 시기에 따른 분포 양상이 유사하였고, 실제 신규 확진자 발생의 분포와도 비교하였을 때도 Ct value의 분포에 따른 확진자 발생 예측이 가능하다는 것을 확인하였다.

이를 바탕으로 실제 Ct value를 통한 신규 확진자 발생 예측 분석을 통하여 국가 단위 검사실 기반 감시체계 활용 가능성을 보고자 하였다. 예측 분석을 위하여 먼저 단순 선형회귀모형을 통해 Ct value에 time lag(1일부터 14일 별)를 반영하여 신규 확진 수를 예측했다.

전체 시기(2020. 02. 18. - 2021. 12. 31., 총 668일)를 대상으로 분석을 진행한 결과, 모든 lag에서 유의하였고 그 중 11일 lag(시차)에서 가장 강한 예측 설명력을 보였다(Figure 6, Table 3). 11일 lag 이전에는 점차 예측 설명력이 증가하다가, 11일 lag를 기점으로 이후부터는 예측 설명력이 점차 감소하였다.

제1기 (2020. 02. 18. - 2020. 08. 11., 총 160일)를 대상으로 분석을 진행한 결과, 당일(lag0)을 제외한 모든 lag에서 유의하였고 그 중 9일 lag에서 가장 강한 예측 설명력을 보였다(Figure 7, Table 4). 9일 lag 이전에는 점차 예측 설명력이 증가하다가, 9일 lag를 기점으로 이후부터는 예측 설명력이 점차 감소하였다.

제2기 (2020. 08. 12. - 2020. 11. 12., 총 79일)를 대상으로 분석을 진행한 결과, 당일(lag0)을 제외한 모든 lag에서 유의하였고 그 중 5일 lag에서 가장 강한 예측 설명력을 보였다(Figure 8, Table 5). 5일 lag 이전에는 점차 예측 설명력이 증가하다가, 5일 lag를 기점으로 이후부터는 예측 설명력이 감소하는 추이를 보였다(8일 lag 제외).

제3기 (2020. 11. 13. - 2021. 07. 06., 총 222일)를 대상으로 분석을 진행한 결과, 1일과 2일 lag를 제외한 모든 lag에서 유의하였고 그 중 14일 lag에서 가장 강한 예측 설명력을 보였다(Figure 9, Table 6). 14일 lag까지 지속적으로 예측 설명력이 증가하는(8일 lag 제외) 양상을 보였다.

제4기 (2021. 07. 07. - 2021. 12. 31. 총 164일)를 대상으로 분석을 진행한 결과, 1일과 2일 lag를 제외한 모든 lag에서 유의하였고 그 중 10일 lag에서 가장 강한 예측 설명력을 보였다(Figure 10, Table 7). 그러나 3일 lag 이전에는 점차 예측 설명력이 증가하다가, 3일 lag를 기점으로 이후부터는 예측 설명력이 감소하는 추이를 보였다(10일 lag 제외).

해당 결과는 정적 추정의 결과로, 동시대적인 관계만을 파악하여 단순히 어떤 시차의 Ct value가 신규 확진 수를 가장 잘 예측하는지에 대한 결과이다. 그러나 시계열 자료의 특성 상 과거의 Ct value 값이 미래의 Ct value 값에 영향을 줄 수 있다고 접근하는 동적 추정의 결과 또한 제시할 필요가 있다.

Table 4. Simple Linear regression results on number of newly confirmed cases by lagged Ct value RdRp (by date, all periods)

	lag0	lag1	lag2	lag3	lag4	lag5	lag6	lag7	lag8	lag9	lag10	lag11	lag12	lag13	lag14
Coefficient (t)	- 0.344*** (- 27.54)	- 0.351*** (- 29.47)	- 0.357*** (- 30.61)	- 0.358*** (- 30.95)	- 0.360*** (- 31.38)	- 0.363*** (- 31.93)	- 0.364*** (- 31.94)	- 0.365*** (- 32.31)	- 0.369*** (- 33.00)	- 0.370*** (- 33.31)	- 0.371*** (- 33.39)	- 0.371*** (- 33.47)	- 0.371*** (- 33.34)	- 0.370*** (- 33.03)	-0.371*** (-33.02)
Constant (t)	13.83*** (46.98)	14.02*** (49.83)	14.15*** (51.46)	14.19*** (51.95)	14.25*** (52.55)	14.31*** (53.29)	14.32*** (53.29)	14.37*** (53.79)	14.45*** (54.75)	14.49*** (55.16)	14.50*** (55.25)	14.52*** (55.33)	14.51*** (55.12)	14.49*** (54.66)	14.50*** (54.62)
Observationis	683	682	681	680	679	678	677	676	675	674	673	672	671	670	669
Adjusted R-squared	0.526	0.56	0.579	0.585	0.592	0.601	0.601	0.607	0.618	0.622	0.624	0.625	0.624	0.620	0.620

※Time lag on Ct value RdRp median
※ t statistics in parentheses
※ * p<0.05, ** p<0.01, *** p<0.001

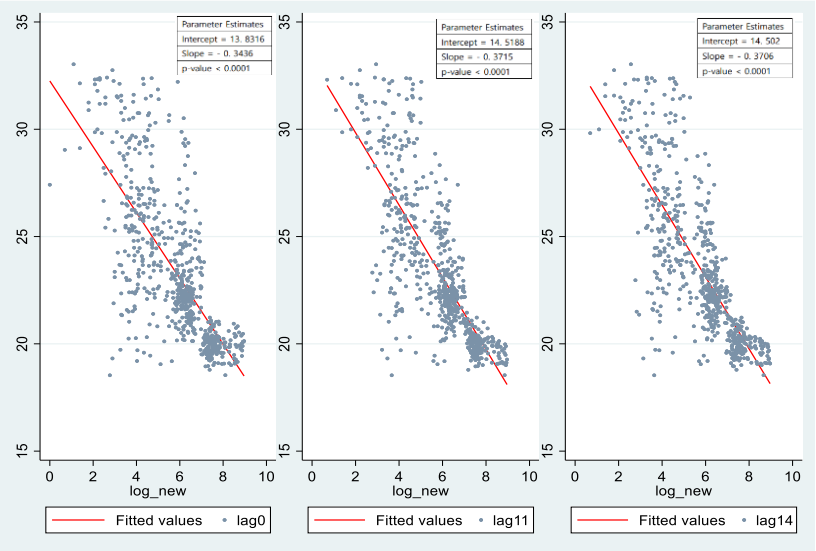


Figure 6. Simple Linear regression graph on number of newly confirmed cases by lagged Ct value RdRp (by date, all periods)

Table 5. Simple Linear regression results on number of newly confirmed cases by lagged Ct value RdRp (by date, 1st periods)

	lag0	lag1	lag2	lag3	lag4	lag5	lag6	lag7	lag8	lag9	lag10	lag11	lag12	lag13	lag14
Coefficient (t)	- 0.0193 (- 0.82)	- 0.0461* (- 2.04)	- 0.0561* (- 2.50)	- 0.0560* (- 2.48)	- 0.0621** (- 2.76)	- 0.0701** (- 3.14)	- 0.0712** (- 3.20)	- 0.0780*** (- 3.55)	- 0.0849*** (- 3.89)	- 0.0865*** (- 3.98)	- 0.0845*** (- 3.95)	- 0.0857*** (- 4.08)	- 0.0826*** (- 4.02)	- 0.0819*** (- 4.07)	-0.0834*** (- 4.23)
Constant (t)	4.309*** (6.63)	5.065*** (8.12)	5.345*** (8.61)	5.339*** (8.57)	5.504*** (8.86)	5.712*** (9.28)	5.734*** (9.34)	5.912*** (9.73)	6.095*** (10.11)	6.124*** (10.22)	6.055*** (10.24)	6.070*** (10.47)	5.968*** (10.53)	5.932*** (10.68)	5.958*** (54.62)
Observationis	176	175	174	173	172	171	170	169	168	167	166	165	164	163	162
Adjusted R-squared	- 0.002	0.018	0.029	0.029	0.037	0.050	0.052	0.064	0.078	0.082	0.624	0.625	0.624	0.620	0.620

※Time lag on Ct value RdRp median
※ t statistics in parentheses
※ * p<0.05, ** p<0.01, *** p<0.001

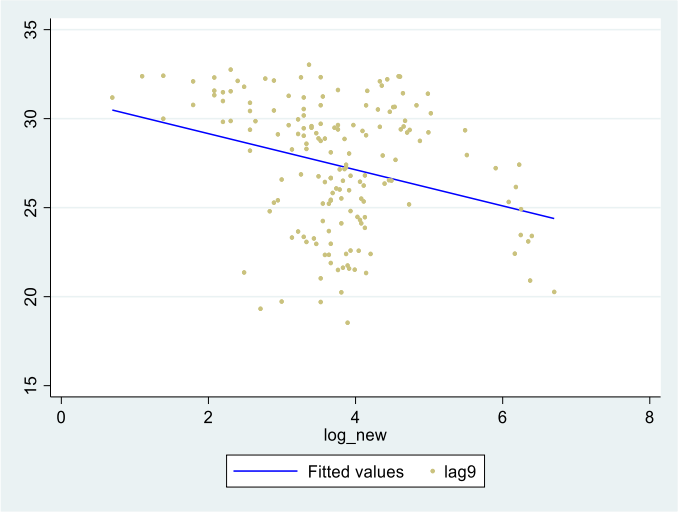


Figure 7. Simple Linear regression graph on number of newly confirmed cases by lagged Ct value RdRp (by date, 1st periods)

Table 6. Simple Linear regression results on number of newly confirmed cases by lagged Ct value RdRp (by date, 2nd periods)

	lag0	lag1	lag2	lag3	lag4	lag5	lag6	lag7	lag8	lag9	lag10	lag11	lag12	lag13	lag14
Coefficient (t)	- 0.0583* (- 2.08)	- 0.0978*** (- 3.61)	- 0.131*** (- 5.20)	- 0.138*** (- 5.75)	- 0.140*** (- 5.94)	- 0.144*** (- 6.19)	- 0.126*** (- 5.23)	- 0.113*** (- 4.56)	- 0.128*** (- 5.40)	- 0.121*** (- 5.29)	- 0.106*** (- 4.57)	- 0.0980*** (- 4.26)	- 0.0924*** (- 3.98)	- 0.0649** (- 2.73)	- 0.0531* (- 2.21)
Constant (t)	6.203*** (9.18)	7.159*** (10.93)	7.968*** (13.04)	8.144***(13.96)	8.186*** (14.32)	8.288*** (14.67)	7.863*** (13.37)	7.559*** (12.46)	7.927*** (13.63)	7.753*** (13.83)	7.392*** (12.98)	7.211*** (12.68)	7.077*** (12.33)	6.404*** (10.85)	6.114*** (10.25)
Observations	93	93	93	93	93	93	93	93	93	93	93	93	93	93	93
Adjusted R-squared	0.035	0.116	0.221	0.258	0.272	0.289	0.222	0.177	0.234	0.227	0.178	0.157	0.139	0.065	0.041

※Time lag on Ct value RdRp median
※ t statistics in parentheses
※ * p<0.05, ** p<0.01, *** p<0.001

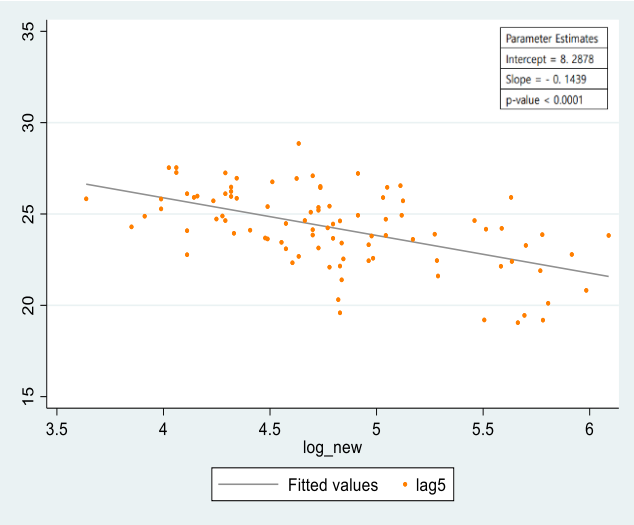


Figure 8. Simple Linear regression graph on number of newly confirmed cases by lagged Ct value RdRp (by date, 2nd periods)

Table 7. Simple Linear regression results on number of newly confirmed cases by lagged Ct value RdRp (by date, 3rd periods)

	lag0	lag1	lag2	lag3	lag4	lag5	lag6	lag7	lag8	lag9	lag10	lag11	lag12	lag13	lag14
Coefficient (t)	- 0.0104* (- 0.71)	- 0.0158 (- 1.07)	- 0.0284 (- 1.92)	- 0.0372* (- 2.52)	- 0.0465** (- 3.18)	- 0.0527*** (- 3.63)	- 0.0578*** (- 4.00)	- 0.0593*** (- 4.09)	- 0.0614*** (- 4.22)	- 0.0605*** (- 4.14)	- 0.0706*** (- 4.89)	- 0.0712*** (- 4.92)	- 0.0726*** (- 5.01)	- 0.0768*** (- 5.34)	- 0.0806*** (- 5.65)
Constant (t)	6.518*** (19.46)	6.639*** (19.83)	6.924*** (20.68)	7.125*** (21.26)	7.335*** (22.07)	7.477*** (22.68)	7.593*** (23.10)	7.628*** (23.23)	7.655*** (23.02)	7.753*** (13.83)	7.886*** (23.98)	7.900*** (23.96)	7.931*** (24.05)	8.027*** (24.50)	8.115*** (24.97)
Observations	236	236	236	236	236	236	236	236	236	236	236	236	236	236	236
Adjusted R-squared	- 0.002	0.001	0.011	0.022	0.037	0.049	0.060	0.063	0.067	0.064	0.089	0.090	0.093	0.105	0.116

※Time lag on Ct value RdRp median
※ t statistics in parentheses
※ * p<0.05, ** p<0.01, *** p<0.001

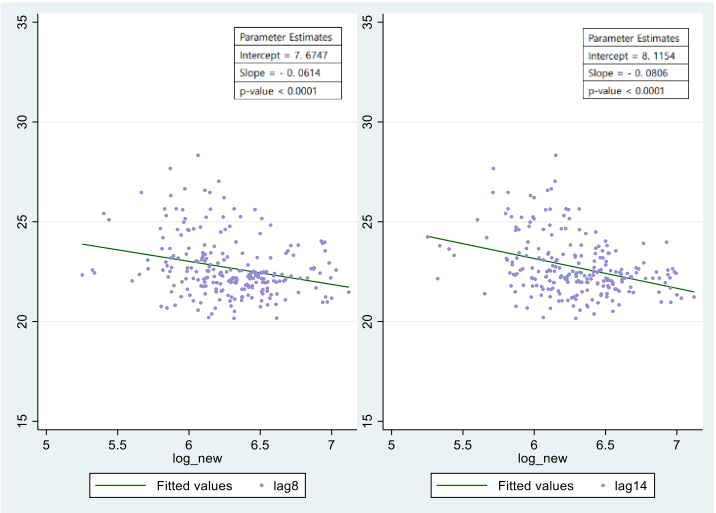


Figure 9. Simple Linear regression graph on number of newly confirmed cases by lagged Ct value RdRp (by date, 3rd periods)

Table 8. Simple Linear regression results on number of newly confirmed cases by lagged Ct value RdRp (by date, 4th periods)

	lag0	lag1	lag2	lag3	lag4	lag5	lag6	lag7	lag8	lag9	lag10	lag11	lag12	lag13	lag14
Coefficient (t)	- 0.0519 (- 0.70)	- 0.0790 (- 1.06)	- 0.227** (- 3.11)	- 0.338*** (- 4.74)	- 0.326*** (- 4.55)	- 0.281*** (- 3.88)	- 0.282*** (- 3.95)	- 0.259*** (- 3.59)	- 0.258*** (- 3.57)	- 0.380*** (- 5.49)	- 0.464*** (- 7.11)	- 0.432*** (- 6.56)	- 0.371*** (- 5.55)	- 0.321*** (- 4.97)	- 0.273*** (- 4.32)
Constant (t)	8.761*** (5.90)	9.302*** (6.23)	12.26*** (8.40)	14.47*** (10.18)	14.23*** (9.95)	13.34*** (9.22)	13.36*** (9.36)	12.90*** (8.95)	12.87*** (8.92)	15.31*** (11.08)	16.98*** (13.04)	16.35*** (12.43)	15.14*** (11.33)	14.14*** (10.95)	13.19*** (10.42)
Observations	178	178	178	178	178	178	178	178	178	178	178	178	178	178	178
Adjusted R-squared	- 0.003	0.001	0.047	0.108	0.100	0.074	0.076	0.063	0.062	0.141	0.219	0.192	0.144	0.118	0.091

※Time lag on Ct value RdRp median
※ t statistics in parentheses
※ * p<0.05, ** p<0.01, *** p<0.001

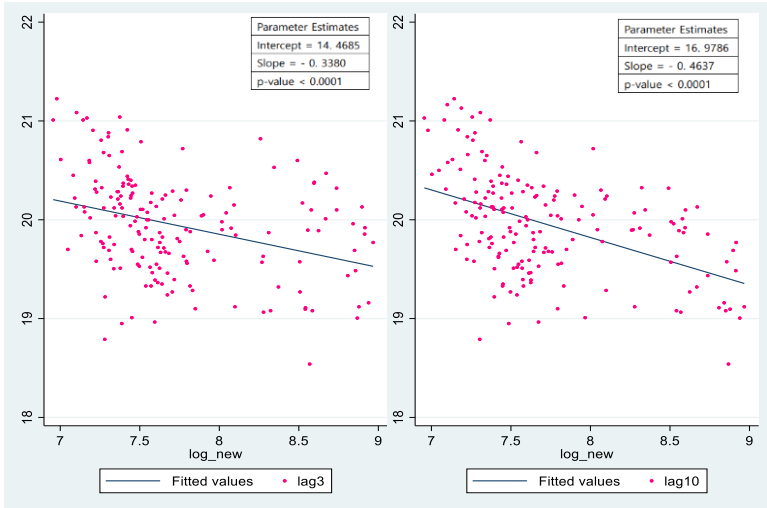


Figure 10. Simple Linear regression graph on number of newly confirmed cases by lagged Ct value RdRp (by date, 4th periods)

3.3.2. Dynamic model: Distributed Lag Model (DLM)

앞서 단순선형회귀 결과는 시계열의 정상성을 가정하지 않은 정적 추정의 회귀 결과로, 동시대적인 관계만을 파악하여 단순히 어떤 시차의 Ct value가 신규 확진 수를 가장 잘 예측하는지에 대한 결과이기에 허위회귀(spurious regression)일 가능성이 있다. 특히 과거의 Ct value 값은 viral load와 연관성이 높고 이는 전파력과도 이어지므로, 신규 확진 수는 단순히 동시대적인 정적 추정의 결과로 예측하기 어렵다. 따라서 과거의 Ct value 값이 미래의 Ct value 값에 영향을 줄 수 있는 자료의 특성(시계열)을 반영한 동적 추정의 결과를 제시하고자 시계열 분석 기법 중 하나인 시차분포모형(DML)에 14일 시차를 준 유한시차분포모형(FDL model)을 활용하였다.

모형 적용 전 사전분석으로, 독립변수와 종속변수의 회귀모형을 추정 후, 독립변수와 종속변수 각각의 과거 값에 대한 단위근(Unit root) 검정을 수행하고 정상 시계열로 변환하였다(Appendix C).

전체 기간을 대상으로 먼저 분석을 한 후, 코로나-19 유행 시기(제1, 2, 3, 4기)별로 각각 분석하여 시간적 변동을 제거한 Ct value를 통한 신규 확진 수를 예측하였다. 효과추정치를 모든 시차에 대해 구한 후, 가장 유의한 시차를 도출하였다. 모델 적용 후 자기상관 검정 결과는 Appendix D에서 확인할 수 있다.

Table 9는 시간적 변동에 의한 효과를 제거하고 본 각 시차에 대한 회귀 계수(Coefficients)를, Table 10은 14일 장기효과추정치를, Figure 11에서는 시기별 유의한 효과추정치를 보인 시차에 대해 나타내었다.

시간적 변동에 의한 효과를 제거한 후 단순히 각 시차에서 회귀 계수를 확인한 결과, 전체 시기와 제1기에서는 1일 lag에서 가장 유의하였고 제2기, 제3기에서는 모든 시차에서 유의하지 않았으며, 제4기에서는 2일 lag에서 가장 유의하였다(Table 9). 이 결과는 앞서 단순선형회귀모형의 결과에서 정상 시계열만 가정한 결과로 정적 추정에 해당한다. 정적 추정이라는 점은 동일하나, 단순선형회귀모형의 결과는 5-14일의 시차에서 가장 잘 예측되었다는 점에서 시차가 비교적 더 길었다.

동적 추정으로 접근하여, 최대 잠복기를 반영한 14일 장기효과추정치는 모든 유행 시기에서 유의하지 않았다(Table 10). 각

유행시기별로 언제 가장 유의한 예측 효과추정치를 보이는지 계산한 결과, 전체 시기에서는 2일 lag까지, 제1기는 1일 lag까지, 제2기는 모든 lag에서 유의하지 않았고, 제3기와 제4기는 모두 3일 lag까지 가장 유의한 예측 효과추정치를 보였다(Figure 11). 제2기에서 유의하지 않은 결과가 나온 것은 제 2기에 해당하는 기간이 너무 짧아서(n = 79)일 가능성이 있다.

사후분석으로 예측 모델이 잘 적합되었는지 확인하기 위해 자기상관 검정을 수행한 결과(Appendix D), 전체 시기, 제1기와 2기는 귀무가설(잔차 간 자기상관이 없다)을 기각하여 자기상관이 있었고, 제3기와 4기는 귀무가설을 기각하지 못하여 자기상관이 없었다. 그러므로 시계열 예측 결과가 제3기와 4기에서 가장 적합하다고 판단하였다.

따라서 제3기와 4기에서 효과추정치는 두 시기 모두 3일 lag까지가 가장 유의하였으므로 Ct value가 3일 lag까지 반영되었을 때 신규 확진수를 가장 잘 예측한다고 할 수 있다.

Table 3. Finite Distribution lag model of the number of newly confirmed cases by Ct value RdRp (by date)

	Total periods	1st Periods	2nd Periods	3rd Periods	4th Periods
lag0	- 0.0167* (0.01)	- 0.0331* (0.01)	- 0.0166 (0.03)	0.0065 (0.01)	0.0386 (0.03)
lag1	- 0.0318*** (0.01)	- 0.0471** (0.01)	0.0093 (0.03)	- 0.0087 (0.01)	0.0483 (0.03)
lag2	- 0.0041 (0.01)	0.0212 (0.02)	- 0.0223 (0.03)	- 0.0141 (0.01)	- 0.1525*** (0.03)
lag3	0.0232** (0.01)	0.0463** (0.02)	- 0.0097 (0.03)	- 0.0209 (0.01)	- 0.0519 (0.03)
lag4	0.0062 (0.01)	- 0.0046 (0.02)	0.0056 (0.03)	0.0005 (0.01)	0.0658* (0.03)
lag5	- 0.0016 (0.01)	- 0.0106 (0.02)	0.0038 (0.03)	0.0083 (0.01)	0.0393 (0.03)
lag6	0.0180* (0.01)	0.0215 (0.02)	0.0350 (0.03)	0.0021 (0.01)	- 0.0468 (0.03)
lag7	0.0072 (0.01)	0.0010 (0.02)	0.0130 (0.03)	0.0173 (0.01)	0.0438 (0.03)
lag8	- 0.0137 (0.01)	- 0.0163 (0.02)	- 0.0468 (0.03)	0.0150 (0.01)	0.0679* (0.03)
lag9	- 0.0061 (0.01)	- 0.0000 (0.02)	0.0157 (0.03)	0.0069 (0.01)	- 0.1201*** (0.03)
lag10	- 0.0013 (0.01)	0.0027 (0.02)	- 0.0078 (0.03)	- 0.0258 (0.01)	- 0.0417 (0.03)
lag11	- 0.0019 (0.01)	- 0.0112 (0.02)	- 0.0027 (0.03)	0.0007 (0.01)	0.0495 (0.03)
lag12	0.0091 (0.01)	0.0180 (0.02)	- 0.0465 (0.03)	0.0062 (0.01)	0.0388 (0.03)
lag13	0.0117 (0.01)	0.018 (0.02)	0.0473 (0.03)	- 0.0194 (0.01)	0.0150 (0.03)
lag14	- 0.0032 (0.01)	- 0.0122 (0.01)	0.0091 (0.03)	0.0173 (0.01)	0.0048 (0.03)
_cons	0.1158 (0.08)	0.3046 (0.33)	0.3293 (0.88)	0.1870 (0.21)	0.0292 (0.68)
Observations	668	160	79	222	164
Adjusted R-squared	0.0549	0.0577	- 0.0495	0.003	0.4974

※Time lag on Ct value RdRp median
※ Coefficients in columns
※ Standard error in parentheses
※ * p<0.05, ** p<0.01, *** p<0.001

Table 10. T-test results of linear combinations of estimators

	Coefficient	Standard error	t	P > |t|	[95% confidence interval]
Total periods	- 0.0049	0.0032	- 1.52	0.129	[- 0.0113, 0.0014]
1st Periods	- 0.0117	0.0120	- 0.98	0.329	[- 0.0353, 0.0119]
2nd Periods	- 0.0136	0.0361	- 0.38	0.707	[- 0.0857, 0.0584]
3rd Periods	- 0.0082	0.0092	- 0.90	0.371	[- 0.0264, 0.0099]
4th Periods	- 0.0010	0.0340	- 0.03	0.977	[- 0.0681, 0.0661]

※ Long Run Propensity(LPR) of Ct value RdRp median
※ Linear combinations t-test of estimation coefficients:
lag0 + lag1 + lag2 + lag3 + lag4 + lag5 + lag6 + lag7 + lag8 + lag9 + lag10 + lag11 + lag12 + lag13 + lag14 = 0

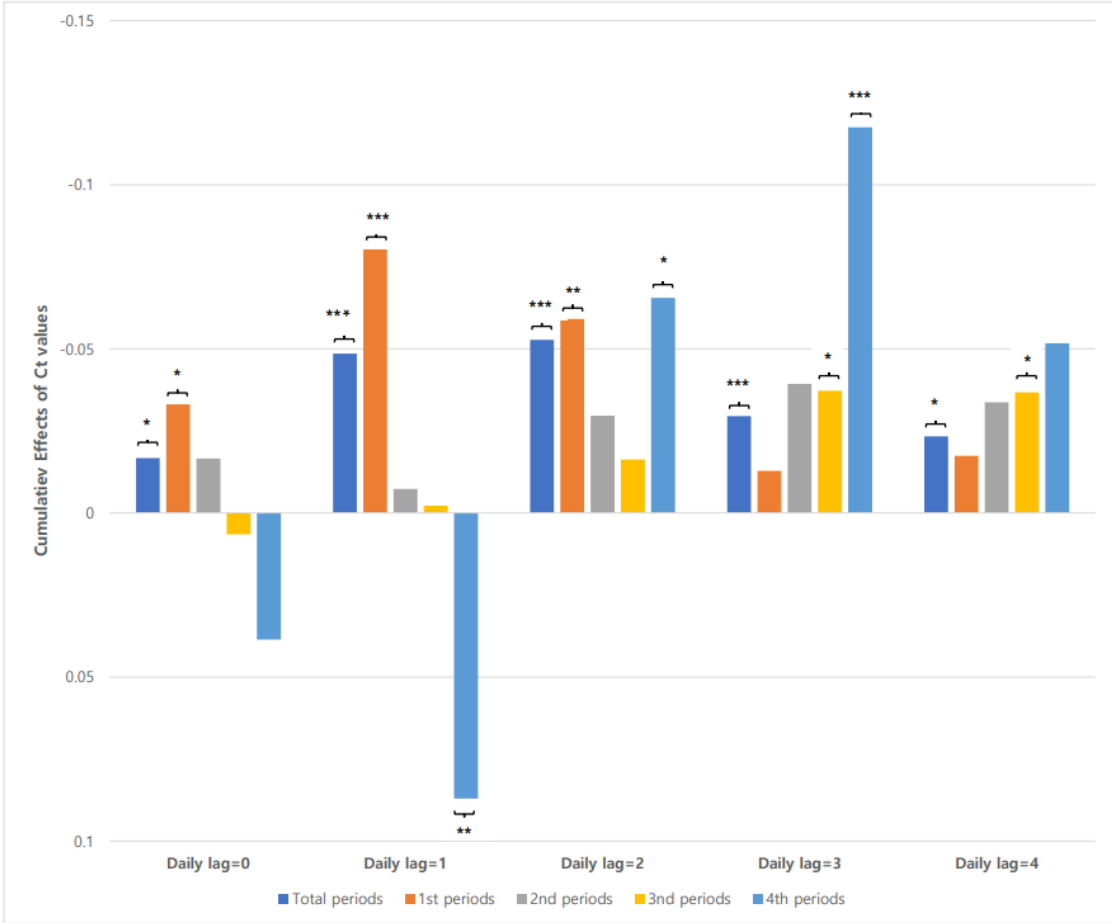


Figure 11. LPR of Ct values on newly confirmed cases

Forecasting the number of newly confirmed cases based on the Reproduction rate(Rt)

Ct values를 활용한 예측과 비교를 위해 감염재생산지수(Rt) 또한 동적 추정의 결과를 제시하고자 동일하게 14일 시차를 반영한 유한시차분포모형(FDL model)을 활용하였다. 모형 적용 전 사전분석으로 단위근(Unit root) 검정 및 정상 시계열 변환 과정을 수행하였다.

전체 기간 및 코로나-19 유행 시기(제1, 2, 3, 4기)별로 각각 분석하여 시간적 변동을 제거한 Rt값으로 신규 확진 수를 예측하였다. 효과추정치를 모든 시차에 대해 구한 후, 가장 유의한 시차를 도출하였다. 모델 적용 후, 자기상관 검정을 수행하여 분석 결과를 신뢰할 수 있는지 확인하였다.

시간적 변동에 의한 효과를 제거한 후 단순히 각 시차에서 회귀계수를 확인한 결과, 모든 시기의 1일, 2일 lag에서 가장 유의하였고 4일 lag부터는 유의하지 않았다(Table 11). Ct values는 제4기에서 회귀모형의 설명력(Adjusted R-squared)이 가장 높았고 제2기에서 가장 낮았는데, Rt 는 반대로 제 2기에서 회귀모형의 설명력이 가장 높았고 제4기에서 가장 낮았다. 이는 앞서 분포를 비교한 결과인 “Rt가 감염 경로가 파악되지 않는 경우가 증가되고 확진 수가 급증한 시기인 제3기, 4기에서 값의 변동이 심하지 않아 확진 양상을 제대로 반영하지 못하고 있다”는 것을 더욱 뒷받침한다.

동적 추정으로 접근하여, 각 유행시기별로 언제 가장 유의한 예측 효과추정치를 보이는지 계산한 결과, 모든 시기에서는 1일 lag까지, 가장 유의한 예측 효과추정치를 보였다(Figure 12). 사후분석으로 자기상관이 존재하는 시기는 전체 시기와 제1기였으며, 제2, 3, 4기에서는 자기상관이 존재하지 않았다. 그러므로 시계열 예측 결과가 제2, 3, 4기에서 가장 적합하다고 판단할 수 있다. Rt 값 자체가 일주일의 기간을 두고 계산된 것이므로 Rt는 실질적으로 8일 lag를 두고 신규 확진 수를 가장 잘 예측한다고 할 수 있다.

Table 11. Finite Distribution lag model of the number of newly confirmed cases by Rt (by date)

	Total periods	1st Periods	2nd Periods	3rd Periods	4th Periods
lag0	- 0.9620*** (0.11)	- 0.6698** (0.24)	- 1.7186*** (0.33)	- 1.5298*** (0.26)	- 1.0773* (0.47)
lag1	3.5112*** (0.22)	2.8834*** (0.46)	5.6005*** (0.52)	4.9134*** (0.48)	3.9528*** (0.93)
lag2	- 3.3229*** (0.24)	- 2.7363*** (0.50)	- 5.2025*** (0.57)	- 4.4580*** (0.50)	- 3.7170*** (1.05)
lag3	0.9851*** (0.24)	0.7999 (0.50)	1.6604** (0.57)	1.2846** (0.49)	0.8331 (1.06)
lag4	- 0.1256 (0.24)	- 0.3491 (0.49)	0.0837 (0.57)	- 0.1876 (0.48)	0.5813 (1.06)
lag5	0.1899 (0.23)	0.5235 (0.48)	- 0.6621 (0.57)	0.3260 (0.48)	- 0.5575 (1.06)
lag6	- 0.1673 (0.23)	- 0.3181 (0.47)	0.2892 (0.56)	- 0.3813 (0.47)	- 0.0282 (1.04)
lag7	- 0.0156 (0.22)	- 0.0620 (0.46)	0.1029 (0.57)	0.1471 (0.46)	0.2133 (1.02)
lag8	0.1346 (0.22)	0.0817 (0.48)	0.1224 (0.60)	0.0759 (0.47)	- 0.0274 (1.04)
lag9	- 0.2712 (0.22)	- 0.1520 (0.48)	- 0.2449 (0.61)	- 0.1236 (0.47)	- 0.0956 (1.06)
lag10	0.3410 (0.23)	0.2796 (0.48)	0.3935 (0.59)	0.2222 (0.47)	- 0.0892 (1.06)
lag11	- 0.1242 (0.23)	- 0.0238 (0.48)	0.1089 (0.50)	- 0.3287 (0.47)	0.3691 (1.06)
lag12	- 0.2045 (0.23)	- 0.3711 (0.48)	- 0.6766 (0.47)	0.2207 (0.47)	- 0.3157 (1.05)
lag13	0.3040 (0.20)	0.4707 (0.41)	0.4889 (0.41)	- 0.0118 (0.43)	0.1168 (0.93)
lag14	- 0.1504 (0.09)	- 0.2222 (0.19)	- 0.1525 (0.21)	- 0.0376 (0.24)	- 0.0052 (0.46)
_cons	- 0.1220** (0.04)	- 0.1348 (0.09)	- 0.1839 (0.12)	- 0.1319* (0.06)	- 0.1534 (0.13)
Observations	665	158	79	222	164
Adjusted R-squared	0.4028	0.3348	0.6834	0.3832	0.0094

※Time lag on Reproduction rate(Rt)
※ Coefficients in columns
※ Standard error in parentheses
※ * p<0.05, ** p<0.01, *** p<0.001

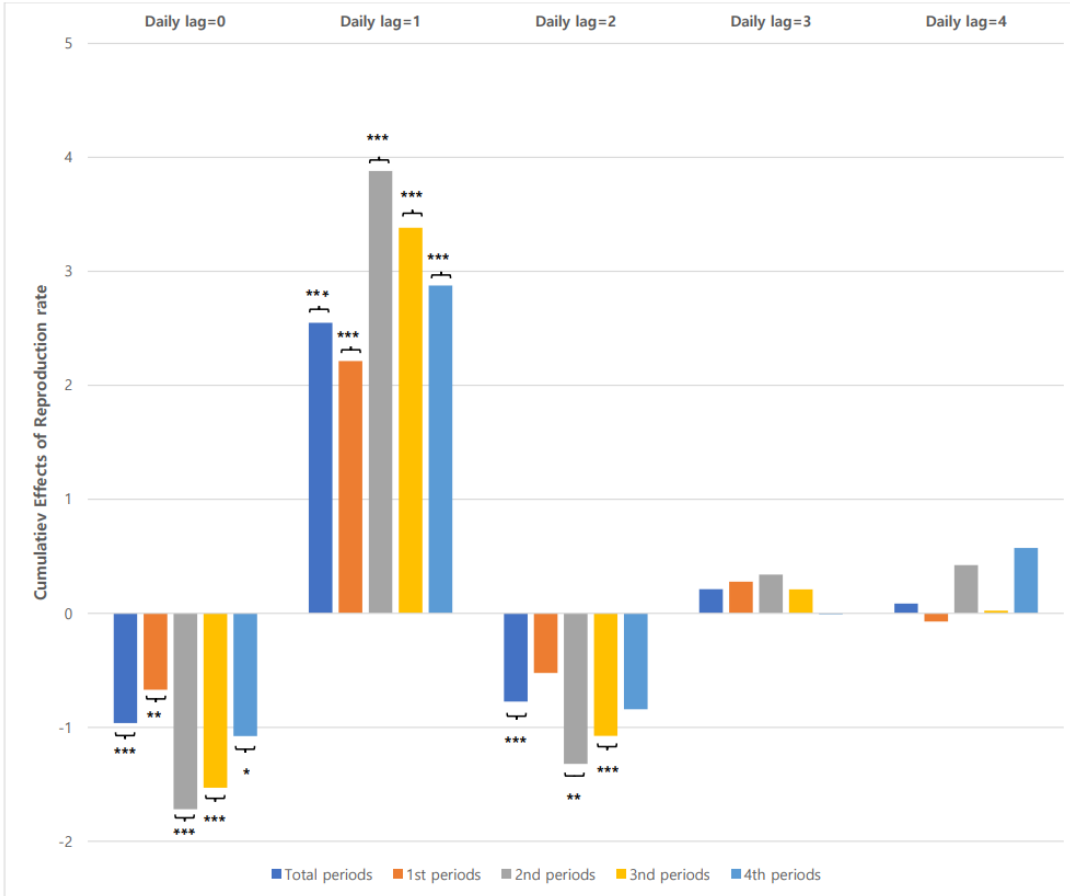


Figure 12. LPR of Rt on newly confirmed cases

Forecasting the number of newly confirmed cases in the time of Omicron variants spread

코로나-19의 변종 바이러스 중 오미크론 변이는 2021년 12월 1일에 국내에 처음 등장하였다. 오미크론은 델타 변이에 비해 감염이 쉽고 전파력이 높으며, 중증도가 낮다는 특징이 있어 등장 이후 신규 확진자 수가 큰 폭으로 증가하였다. 본 연구에서는 오미크론 확산 시기를 특정하여 Rt와 Ct values를 통한 동적 추정의 결과를 서로 비교하여 변이 바이러스 시기에서의 예측력을 파악하고자 하였다(Figure 13, 14). 그러나 자료의 최대 기간이 2021년 12월 31일까지로 분석 대상 기간이 30일에 불과하게 되어 유의한 예측결과를 확인하기 어려웠다. 또한, 오미크론이 국내 우세종이 된 시기는 2022년 1월 24일부터로(질병관리청, 2022), 2021년 12월 한 달동안에는 정확한 오미크론의 확산세를 반영하지 못하였다.

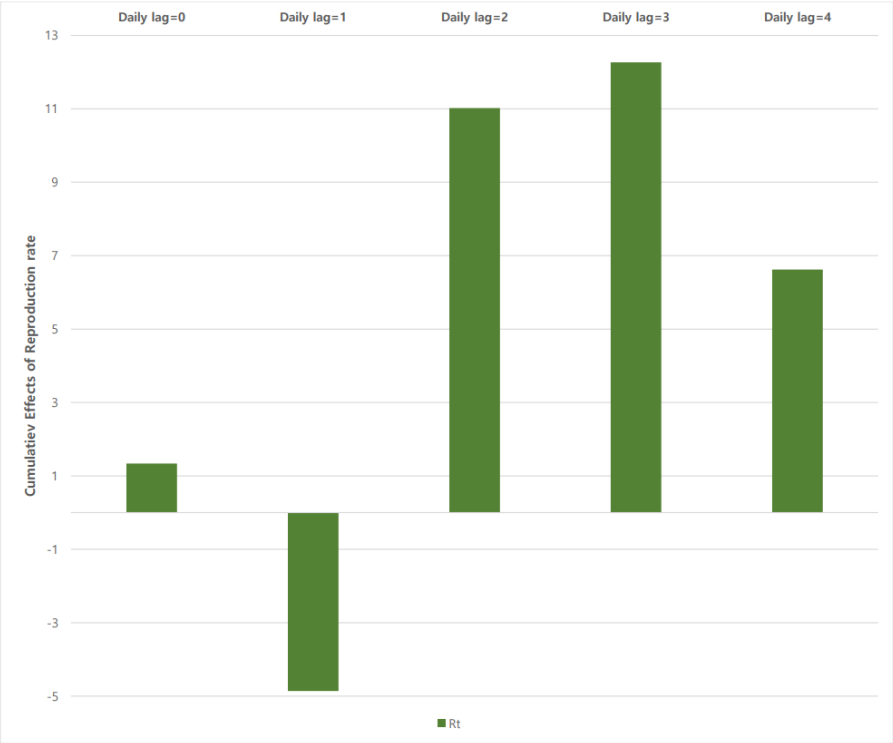


Figure 13. LPR of Rt on newly confirmed cases in the time of Omicron variants spread

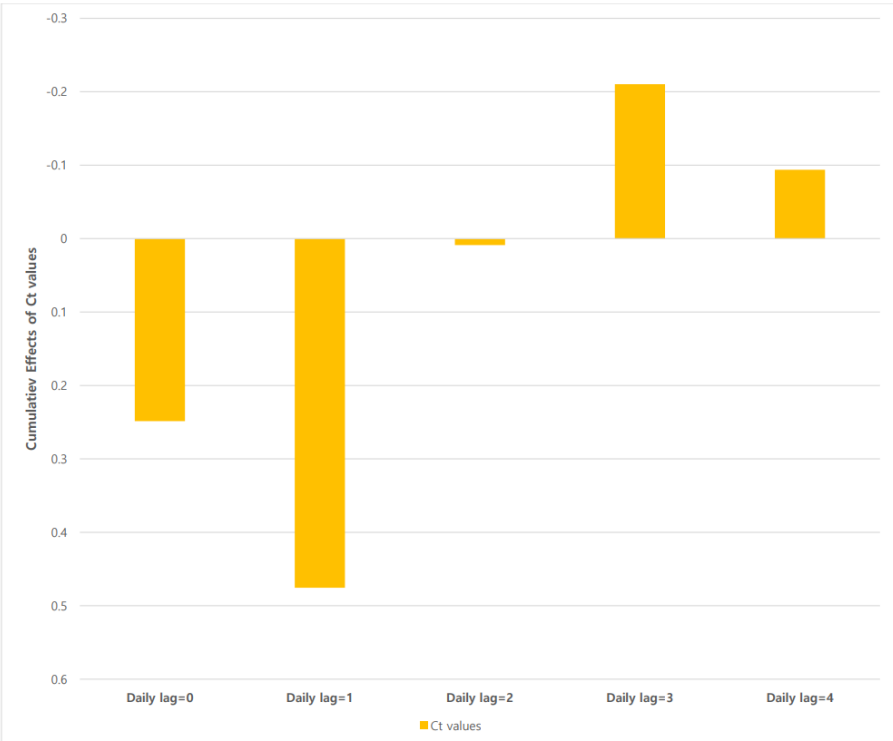


Figure 14. LPR of Ct values on newly confirmed cases in the time of Omicron variants spread

4. DISCUSSION

4.1. Summary of results

본 연구는 인구집단 수준의 국내 데이터로는 거의 최초로 Ct value 분포를 파악하고 신규 확진 수를 예측한 연구로, Ct value의 보건정책 수립 근거로 활용 가능성을 제시하고자 하였다. 진단검사 결과의 각 변수에 대한 기술통계 표 및 그래프를 제시하고, 코로나-19 유행 시기별 인구집단 수준의 Ct value 분포 양상을 시각적으로 확인한 후 실제 Ct value를 통한 신규 확진 발생 예측 분석을 두 가지 방법(단순 선형회귀모형, FDL model)으로 수행하였다.

기술통계 결과, 진단검사 대부분이 검사전문기관에서 이루어졌고, 20-50대의 확진 환자가 전체 확진 환자의 절반 이상이었다. 씨젠 진단 키트 제품이 절반 이상 사용되었고, 거의 모든 검사에서 동일한 유전자 증폭장비를 사용하였으며, 검체로는 인후도말(NPS/OPS)이 절반 이상 사용되었다. Ct value E, RdRp, N은 모두 비슷한 분포 형태를 보였으며, 값이 10-20 사이에 주로 분포하였다.

두 개의 Ct value 분포 양상 그래프에서 먼저 Rt값의 변동이 Ct values에 비해 신규 확진 수가 급증하는 시기를 충분히 설명하지 못함을 시각적으로 확인하였다. 그 다음 방역정책 변화가 확진 수에 영향을 미치기 이전에 Ct value가 먼저 영향을 받아 값이 변동한다는 것을 확인하였다. 이 두 시계열 분포 그래프를 통해 Ct values의 분포에 따른 확진자 발생 예측이 가능하다고 가정하였다.

Ct value를 통한 실제 신규 확진 발생 예측 분석을 수행하기 위해 정적 추정(단순 선형회귀모형)과 동적 추정(FDL model)을 활용하였다. 먼저 단순 선형회귀모형의 결과 5-14일 시차에서 가장 예측이 잘 되었다. 단순 선형회귀모형을 통한 예측은 예측 설명력이 가장 높은 시차가 시기별로 많이 차이가 났다(전체 시기 11일, 제1기 9일, 제2기 5일, 제3기 14일, 제4기 10일). 그러나 시계열의 정상성을 가정하지 않아 결과가 허위회귀일 가능성이 있고, 신규 확진 수는 단순히 동시대적인 정적 추정의 결과로 예측하기 어려우므로 시간적 변동에 의한 효과를 제거하고 동적 추정을 통한 예측을 수행할 필요가 있었다.

동적 추정 결과 1 - 3일 시차까지에서 가장 예측이 잘 되었고, 유행 시기별로 차이(2일, 1일, 3일, 3일)도 크지 않아 정적 추정의 결과와 차이를 보였다. 사후분석을 통해 단기(3일) 코로나-19 예측지표로 Ct values 값을 활용할 수 있음을 확인하였다. 동일한 분석 방식으로 Rt를 통한 예측의 결과 모든 시기에서 8일 시차에서 가장 예측이 잘 되었다. 오미크론 국내 등장(2021.12.01.)부터 자료의 최대 시기인 2021년 12월 31일까지 기간을 특정하여 오미크론 확산 시기에서의 Ct values와 Rt를 통한 신규 확진 수 예측을 비교하였으나 분석 대상 기간이 30일로 짧았고, 오미크론의 우세종 시기가 2022년 2월부터였으므로 유의한 예측 결과를 확인할 수 없었다.

4.2. Comparison with previous studies

코로나-19 유행 시기별 Ct values 분포를 Rt(감염재생산지수)와 함께 나타낸 시계열 그래프를 통해, Rt가 꾸준히 증가하는 확진 수를 반영하기에 한계를 보이며, 감염 경로가 파악되지 않는 경우가 증가함에 따라 이 한계를 더욱 명확하게 확인하였다. 이는 앞서 선행 연구(유명수 등, 2021; 정재웅 등, 2020)에서 확인한 Rt의 정의 및 계산식 상 감염경로가 파악되지 않으면 정확한 추정이 불가능하다는 결과와 일치하였다.

Ct values를 이용하여 지역사회 유행병 양상 예측 및 감시가 가능하다는 기존 연구(Hay et al., 2021; Andriamandimby et al., 2022; Walker et al., 2021) 결과는 본 연구의 연구목적과 일치하였다.

앞서 Ct values를 활용한 코로나-19 확진 양상 예측(Forecasting)에 관한 연구 11편에 대해 작성한 대조표(Table 2)에 본 연구의 결과를 “Ct values median 값은 3일 lag로 일별 신규 확진 수를 예측할 수 있다” 로 서술하였다. 기존 연구에서 시계열 분석을 활용하지 않고 정적 추정을 통한 Ct value와 신규 확진 수의 연관성 분석 결과는 17일(Yin et al., 2021), 14일(Phillips et al., 2022) 시차에서 가장 밀접하게 연관되므로 열흘 이상의 시차가 존재한다는 점에서 본 연구의 전체 기간 대상 정적 추정 결과인 11일과 유사하였다. 반면 동적 추정을 통한 예측 결과는 Now-casting(초단기예측)(Lin et al., 2022; Khalil et al., 2022), 6-7일(Alizon et al., 2022)에서 가장 잘

예측되므로 7일 이하의 시차를 보인다는 점에서 본 연구의 동적 추정 결과인 3일과 유사하였다. 정적 추정과 동적 추정 모두 기존 연구와 본 연구의 결과가 시차에 차이를 보인 이유는 방법론, 표적 유전자, 연구 기간 및 변이 바이러스 확진 환자 포함 여부가 달랐기 때문일 수 있다. 그 외 코로나-19의 임상적 예후(중증 진행률, 사망률)를 예측한 연구는 주로 병원 및 의료기관에서의 입원 환자를 대상으로 하며 연구기간이 1년 미만이었다. 본 연구에 사용한 자료로는 개인별 임상적 예후를 알 수는 없었으나, 인구집단 수준의 중증화율 및 사망률 지표와 연결하여 분석해 볼 수 있다

4.3. Strengths and Limitations

본 연구의 가장 큰 특장점은 인구집단 수준의 국내 데이터로는 거의 최초로 Ct value의 분포를 파악하고 신규 확진 수 예측에 사용했다는 것이다. 특히 국내 Ct values를 Rt와 비교하며 방역정책의 변화 시기와 함께 살펴보고, 신규 확진 수 예측에 정적 추정과 동적 추정 기법을 모두 활용하는 등 다각적으로 분석한 시도는 이전까지 없었다. 또한 결과적으로 일주일 추정치인 Rt의 한계를 보완할 수 있는 3일 단기 예측지표로 Ct values의 효용을 실제 확인하였다.

그러나 본 연구에는 몇 가지 한계점이 존재한다. 먼저, 개인의 증상여부, 백신접종여부, 변종 바이러스 여부를 알 수 없어 여러 임상적인 특성을 연관 지어 분석할 수 없었다. 한 개인이 아닌 검체에 대한 결과이므로 중복된 개인이 존재한다는 문제점도 있었다. 때문에 분석 시 일별(daily)로 통합하여(aggregate) 중위수 값을 사용하였으나, 여전히 동일 개인이 다른 날에 검사를 받을 경우를 식별할 수 없었다. 또한 전체 검사자의 결과를 사용하여 키트 제조사가 단일로 통일되지 않고 다양하였다. 이는 키트별로 표적유전자와 양성 판독수치에 차이가 존재하는 것 때문에 여러 키트의 Ct values가 aggregated된 값인 daily/weekly median 값을 사용한 분석에 해석에 어려움을 발생시켰다. 마지막으로, 제1기, 제2기, 전체시기를 대상으로 한 동적 추정의 사후분석 결과에서 자기상관 오차의 문제가 존재하여 다중공선성(Multicollinearity)문제가 발생하여 제대로 된 회귀계수를 추정하지 못해 모형 해석에 어려움이 존재하였다.

4.4. Public health implications

본 연구는 향후 국내 코로나-19를 비롯한 신종 바이러스 감염병 확진 양상 예측 시 일주일 추정치인 Rt의 한계를 보완한 단기 예측지표로 Ct value의 활용가능성을 제고할 수 있다는 보건학적 함의를 지닌다. Ct values는 계산이 필요한 추정 지표인 Rt와 달리 PCR을 통한 감염 진단 시에 측정되는 측정 지표이므로, 각 지역 및 기관에서 필요에 따라 손쉽게 예측에 활용할 수 있어 코로나-19를 비롯한 신종 바이러스 감염병 상황 하의 자원 배분과 정책 결정에 신속하게 판단 근거로 사용되어 위기 상황을 예방 및 대응할 수 있다. 특히 대규모 확진 등으로 병상과 인력이 모자라고 방역 단계를 조절할 때, 각 지자체의 상황에 맞게 방역 정책 시행 일정을 유동적으로 조정하거나 인근 지자체로 병상을 이전하는 등 보다 빠른 판단이 필요한 상황에 적극적으로 활용할 수 있을 것으로 생각된다. 또한 점차 가벼운 증상과 높은 확산력을 지닌 인플루엔자와 같이 변화해가는 코로나-19 상황에서는 일상 회복을 위하여 이전에 과도하게 늘린 방역 자원들을 축소 및 유동적으로 운용하는 것이 필요하다. 이 때 어느 규모로, 언제까지 축소하였다가, 또 필요에 따라 여유 자원을 확보할지 의사결정에 Ct values를 활용할 수 있다.

후속 연구에서 증상 여부, 중증도 등의 임상적 특징들과 변이 바이러스 여부, 백신 접종 여부 등이 결합되어 Ct values에도 단계를 구분한 후 각 대상군을 나누어 분석한다면 환자 특성에 맞는 보다 다양한 예측 결과를 제시하여 정책 활용에 Ct values를 다각도로 활용할 수 있을 것이다.

5. CONCLUSION

본 연구는 기존에 국내에서 거의 활용되지 않던 Ct value의 신규 확진 양상 예측이라는 역학적 효용을 확인하였다. 변이 바이러스 등장 시기를 포함한 인구집단 수준의 국내 Ct value 분포를 활용한 코로나-19 예측 연구를 수행하였다. 국내 연구로는 최초로 정적 모델과 동적 모델을 모두 사용하여 Ct value를 통한 예측을 수행한 연구로, Ct value가 단기적(3일) 코로나-19 예측지표로 사용될 수 있음을 밝혔다.

본 연구의 결과를 기반으로 임상적, 정책적 지표들을 결합하여 다양한 후속 연구가 이루어진다면 신규 확진 수 예측뿐만 아니라 Ct value의 역학적 효용이 더욱 확장되어 보건정책 수립의 근거로 확립되는데 기여할 것이다.

REFERENCES

Earnest, R., Uddin, R., Matluk, N., Renzette, N., Turbett, S. E., Siddle, K. J., Loreth, C., Adams, G., Tomkins-Tinch, C. H., Petrone, M. E., Rothman, J. E., Breban, M. I., Koch, R. T., Billig, K., Fauver, J. R., Vogels, C. B. F., Bilguvar, K., De Kumar, B., Landry, M. L., . . . Grubaugh, N. D. (2022). Comparative transmissibility of SARS-CoV-2 variants Delta and Alpha in New England, USA. *Cell Rep Med*, *3*(4), 100583. https://doi.org/10.1016/j.xcrm.2022.100583

Tom, M. R., & Mina, M. J. (2020). To Interpret the SARS-CoV-2 Test, Consider the Cycle Threshold Value. *Clin Infect Dis*, *71*(16), 2252-2254. https://doi.org/10.1093/cid/ciaa619

Aranha, C., Patel, V., Bhor, V., & Gogoi, D. (2021). Cycle threshold values in RT-PCR to determine dynamics of SARS-CoV-2 viral load: An approach to reduce the isolation period for COVID-19 patients. *J Med Virol*, *93*(12), 6794-6797. https://doi.org/10.1002/jmv.27206

Marks, M., Millat-Martinez, P., Ouchi, D., Roberts, C. h., Alemany, A., Corbacho-Monné, M., Ubals, M., Tobias, A., Tebé, C., Ballana, E., Bassat, Q., Baro, B., Vall-Mayans, M., G-Beiras, C., Prat, N., Ara, J., Clotet, B., & Mitjà, O. (2021). Transmission of COVID-19 in 282 clusters in Catalonia, Spain: a cohort study. *The Lancet Infectious Diseases*, *21*(5), 629-636. https://doi.org/10.1016/s1473-3099(20)30985-3

Waudby-West, R., Parcell, B. J., Palmer, C. N. A., Bell, S., Chalmers, J. D., & Siddiqui, M. K. (2021). The association between SARS-CoV-2 RT-PCR cycle threshold and mortality in a community cohort. *Eur Respir J*, *58*(1). https://doi.org/10.1183/13993003.00360-2021

Rodriguez-Grande, C., Catalan, P., Alcala, L., Buenestado-Serrano, S., Adan-Jimenez, J., Rodriguez-Maus, S., Herranz, M., Sicilia, J., Acosta, F., Perez-Lago, L., Munoz, P., Garcia de Viedma, D., & Gregorio Maranon Microbiology, I. D. C. S. G. (2021). Different dynamics of mean SARS-CoV-2 RT-PCR Ct values between the first and second COVID-19 waves in the Madrid population. *Transbound Emerg Dis*, *68*(6), 3103-3106. https://doi.org/10.1111/tbed.14045

이진수. (2020.5.13.). 코로나-19 진단기기 산업현황 및 수출전망(보건산업브리프 vol. 303). 한국보건산업진흥원. 의료기기산업종합지원센터. 보건산업정책연구센터. 권덕철. https://www.khidi.or.kr/board/view?linkId=48827176&menuId=MENU01530

나종화. (2020) R 응용 시계열분석. 자유아카데미.

안선욱, 곽정선, 권수영, 백설화, 배희순, 권근용 등. (2021.7.29.). 세종특별자치시 코로나19 발생의 역학 및 임상적 특성과 진단검사 결과와의 연관성 분석(주간 건강과 질병 제14권 제31호 2229-2239). 질병관리청. 세종특별자치시 감염병관리지원단. https://www.kdca.go.kr/board/board.es?mid=a20602010000&bid=0034&list_no=714174&act=view

Suh, H. J., Kim, D. H., Heo, E. Y., Lee, H. W., Lee, J. K., Lee, C. S., Kim, M., Jeon, Y. D., Chung, J. W., Kim, Y. K., Shin, P. J., Lee, M. S., Kang, J. S., Lee, M. J., Kim, B. N., & Park, S. W. (2020). Clinical Characteristics of COVID-19: Clinical Dynamics of Mild Severe Acute Respiratory Syndrome Coronavirus 2 Infection Detected by Early Active Surveillance. *J Korean Med Sci*, *35*(32), e297. https://doi.org/10.3346/jkms.2020.35.e297

Tizzoni, M., Nash, R. K., Nouvellet, P., & Cori, A. (2022). Real-time estimation of the epidemic reproduction number: Scoping review of the applications and challenges. PLOS Digital Health, 1(6). doi:10.1371/journal.pdig.0000052

Hoque, A., Malek, A., & Zaman, K. (2022). Data analysis and prediction of the COVID-19 outbreak in the first and second waves for top 5 affected countries in the world. Nonlinear Dyn, 109(1), 77-90. doi:10.1007/s11071-022-07473-9

Lin, Y., Yang, B., Cobey, S., Lau, E. H. Y., Adam, D. C., Wong, J. Y., Bond, H. S., Cheung, J. K., Ho, F., Gao, H., Ali, S. T., Leung, N. H. L., Tsang, T. K., Wu, P., Leung, G. M., & Cowling, B. J. (2022). Incorporating temporal distribution of population-level viral load enables real-time estimation of COVID-19 transmission. *Nat Commun*, *13*(1), 1155. https://doi.org/10.1038/s41467-022-28812-9

유명수, 김연주, 백수진, 권동혁. (2021.2.4.). 감염재생산지수 개념 및 방역정책에 따른 변화(주간 건강과 질병 제14권 제6호 282-289). 질병관리청. 질병관리본부 위기대응분석관 위기대응연구담당관. https://www.kdca.go.kr/board/board.es?mid=a20602010000&bid=0034&list_no=712021&act=view

정재웅, 권혁무, 홍성훈 and 이민구. (2020). 국내 코로나바이러스감염증-19의 감염재생산수 추정. 품질경영학회지, 48(3), 493-510.

Hay, J. A., Kennedy-Shaffer, L., Kanjilal, S., Lennon, N. J., Gabriel, S. B., Lipsitch, M., & Mina, M. J. (2021). Estimating epidemiologic dynamics from cross-sectional viral load distributions. *Science*, *373*(6552). https://doi.org/10.1126/science.abh0635

Khalil, A., Al Handawi, K., Mohsen, Z., Abdel Nour, A., Feghali, R., Chamseddine, I., & Kokkolaras, M. (2022). Weekly Nowcasting of New COVID-19 Cases Using Past Viral Load Measurements. *Viruses*, *14*(7).

https://doi.org/10.3390/v14071414

Alizon, S., Selinger, C., Sofonea, M. T., Haim-Boukobza, S., Giannoli, J. M., Ninove, L., Pillet, S., Thibault, V., de Rougemont, A., Tumiotto, C., Solis, M., Stephan, R., Bressollette-Bodin, C., Salmona, M., L'Honneur, A. S., Behillil, S., Lefeuvre, C., Dina, J., Hantz, S., . . . group, S. C.-s. (2022). Epidemiological and clinical insights from SARS-CoV-2 RT-PCR crossing threshold values, France, January to November 2020. *Euro Surveill*, *27*(6). https://doi.org/10.2807/1560-7917.ES.2022.27.6.2100406

양성찬, 장진화, 박신영, 안선희, 김성순, 박수빈 등. (2022.2.17.). 국내 코로나19 확진자 2년 발생 보고서(주간 건강과 질병 제15권 제7호 414-426). 질병관리청. 중앙방역대책본부 역학조사분석단 정보분석팀. https://www.kdca.go.kr/board/board.es?mid=a20602010000&bid=0034&list_no=718709&act=view

Jenkins, G. M., & P., B. G. E. (1976). Time Series Analysis Forecasting and control: Rev. ed. Holden-Day.

Hamilton, J. D. (1994). Time Series Analysis. Princeton, N.J.: Princeton University Press.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., . . . Feng, Z. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med*, *382*(13), 1199-1207. https://doi.org/10.1056/NEJMoa2001316

Hintze, J. L., & Nelson, R. D. (1998). Violin Plots: A Box Plot-Density Trace Synergism. The American Statistician, 52(2), 181–184. https://doi.org/10.2307/2685478

대한진단검사의학회 신종 코로나바이러스 대책위원회, 질병관리본부 감염병분석센터. (2020.2.21.). 코로나바이러스감염증-19 검사실 진단 지침(제2판). https://www.kslm.org/rang_board/list.html?num=16591&code=covid19_press&key=subject_body&keyword=%EA%B2%80%EC%82%AC%EC%8B%A4+%EC%A7%84%EB%8B%A8

대한진단검사의학회 신종 코로나바이러스 대책위원회, 질병관리본부 감염병분석센터. (2021.12.1.). 코로나바이러스감염증-19 검사실 진단 지침(제5판). https://www.kslm.org/rang_board/list.html?num=17518&code=covid19_press&key=subject_body&keyword=5%ED%8C%90

Yin, N., Dellicour, S., Daubie, V., Franco, N., Wautier, M., Faes, C., Van Cauteren, D., Nymark, L., Hens, N., Gilbert, M., Hallin, M., &

Vandenberg, O. (2021). Leveraging of SARS-CoV-2 PCR Cycle Thresholds Values to Forecast COVID-19 Trends. *Front Med (Lausanne)*, *8*, 743988. https://doi.org/10.3389/fmed.2021.743988

Skarzynski, M., McAuley, E. M., Maier, E. J., Fries, A. C., Voss, J. D., & Chapleau, R. R. (2022). SARS-CoV-2 Genome-Based Severity Predictions Correspond to Lower qPCR Values and Higher Viral Load. *Glob Health Epidemiol Genom*, *2022*, 6499217. https://doi.org/10.1155/2022/6499217

Ingberg, E., Ahlstrand, E., Cajander, P., Lof, E., Sundqvist, M., Wegener, M., Liden, M., & Cajander, S. (2022). RT-PCR cycle threshold value in combination with visual scoring of chest computed tomography at hospital admission predicts outcome in COVID-19. *Infect Dis (Lond)*, *54*(6), 431-440. https://doi.org/10.1080/23744235.2022.2035428

Phillips, M. C., Quintero, D., Wald-Dickler, N., Holtom, P., & Butler-Wu, S. M. (2022). SARS-CoV-2 cycle threshold (Ct) values predict future COVID-19 cases. *J Clin Virol*, *150-151*, 105153. https://doi.org/10.1016/j.jcv.2022.105153

Barry, M., & Muayqil, T. (2022). RT-PCR Ct values combined with age predicts invasive mechanical ventilation and mortality in hospitalized COVID-19 patients in a MERS-CoV-endemic country. *Heliyon*, *8*(9), e10525. https://doi.org/10.1016/j.heliyon.2022.e10525

Andriamandimby, S. F., Brook, C. E., Razanajatovo, N., Randriambolamanantsoa, T. H., Rakotondramanga, J. M., Rasambainarivo, F., Raharimanga, V., Razanajatovo, I. M., Mangahasimbola, R., Razafindratsimandresy, R., Randrianarisoa, S., Bernardson, B., Rabarison, J. H., Randrianarisoa, M., Nasolo, F. S., Rabetombosoa, R. M., Ratsimbazafy, A. M., Raharinosy, V., Rabemananjara, A. H., . . . Dussart, P. (2022). Cross-sectional cycle threshold values reflect epidemic dynamics of COVID-19 in Madagascar. *Epidemics*, *38*, 100533. https://doi.org/10.1016/j.epidem.2021.100533

Walker, A. S., Pritchard, E., House, T., Robotham, J. V., Birrell, P. J., Bell, I., Bell, J. I., Newton, J. N., Farrar, J., Diamond, I., Studley, R., Hay, J., Vihta, K. D., Peto, T. E., Stoesser, N., Matthews, P. C., Eyre, D. W., Pouwels, K. B., & team, C.-I. S. (2021). Ct threshold values, a proxy for viral load in community SARS-CoV-2 cases, demonstrate wide variation across populations and over time. *Elife*, *10*. https://doi.org/10.7554/eLife.64683

질병관리청, [1.24.보도참고자료] 오미크론 변이의 특성 분석과 확산 대비(정례브리핑). https://www.kdca.go.kr/board/board.es?mid=a20501010000&bid=0015&list_no=718415&cg_code=&act=view&nPage=78

APPENDIX

Appendix A. Summary of systematic review results

Table A. Summary of COVID-19 surveillance/forecast study using cycle threshold values of RT-PCR

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
Róka, E. 2021	Hungary	Raw sewage samples	Not relevant	Between June, 2020 and October, 2020	44 composite and 21 grab samples	N1	Predict future outbreak	하수를 역학적으로 활용한 데이터가 COVID-19 outbreak 발생을 예측할 수 있는지 확인. 헝가리 부다페스트의 모든 하수처리장에 대해 sewage surveillance를 실시하였고 하수에서 처음으로 포착한 신호는 2주 후 확진자 수 증가를 예측할 수 있었음.
Ade, C. 2021	Germany	Nasopharyngeal swabs	Not relevant	Between March 15th and September 15th 2020	65,878 patients	E, S	Associated with age	환자들의 인구학적 데이터 - 성별, 연령이 Ct value와 어떤 연관성이 있는지 확인. Ct value는 여성이 남성보다 더 높았고, 연령은 80세 이상 집단에서 Ct value가 더 높았으며, 입원환자와 외래환자 간 유의한 차이는 없었음.
Abdulrahman, A. 2021	Kingdom of Bahrain	Nasopharyngeal samples	Not relevant	Between May 4, 2020 and September 30, 2020	63,879 patients	E, RdRp, N	Correlated with the number of daily cases / associated with symptomatic status	인구집단에서 Ct value 값의 분포가 COVID-19 검사 양성률과 어떤 연관성이 있는지 확인하고 Ct value에 영향을 미치는 추가적인 요소 확인. Ct value는 검사양성률과 매우 약한 음의 상관관계를 보였으며 연령과 일일 검사수를 보정한 회귀분석에서 높은 설명력을 보임.
Mahallawi, W. H. 2021	Saudi Arabia	Lower respiratory tract samples, upper respiratory tract samples	Not relevant	Between March 20 and May 22, 2020	3,006 patients	NA	Associated with sex	사우디아라비아의 Madinah city COVID-19 확진자들 대상으로 시행한 연구에서 Ct value와 연관성 있는 인자들을 확인. Ct value는 연령에 따른 유의한 차이는 보이지 않았으며, 남성이 여성보다 더 높은 값을 보였음.
Satlin, M. J.	USA	NA	Not	Between March 15,	6,923	ORF1ab, N2	Correlated with public health	사회적 거리두기 등 공중보건 개입이

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
2021			relevant	2020 and May 14, 2020	patients		interventions, and mortality	COVID-19 확진자 발생에 더해 확진자들의 바이러스량(viral load) 또한 감소시키는지 확인. New York City 6개 병원의 6923명 환자를 대상으로 시행한 연구에서 공중보건학적 개입이 일어난 이후 입원한 환자들의 바이러스량이 감소하는 것을 확인.
Suh, H. J. 2020	Korea	Upper respiratory specimens(naso pharyngeal and throat swabs), lower respiratory specimens(sputum or transtracheal aspirates)	Not relevant	Between February 20, 2020 and April 30, 2020	161 patients	E, RdRp, N	Severity with age, viral clearance with time	국내 초기 COVID-19 확진자들의 감염 초기 임상적 특징과 바이러스 검출기간에 대한 확인. 확진자들을 중상 악화 그룹과 그렇지 않은 그룹으로 나누어 보았을 때 PCR 양성률은 전자에서 22일, 후자에서 32일로 나타남.
Singanayagam, A. 2021	UK	Upper respiratory tract samples	Relevant	ATACCC1(alpha variant): between Sept 13, 2020, and March 13, 2021 / ATACCC2(delta variant): between May 25, and Sept 15, 2021	ATACCC1(alpha variant): 369 / ATACCC2(delta variant): 252	ORF1ab, E	Explain why delta variant being transmitted so effectively in populations with high vaccine coverage	델타 변이는 백신 접종률이 높은 인구집단에서도 전파 속도가 빠름. 델타 변이가 유행하는 지역사회에서 감염된 사람들 중 백신 접종자와 미접종자의 전파 양상과 viral load kinetics를 확인. peak viral load는 백신 접종 상태에 따라 다르지 않았으나, 백신 접종자들이 더 빠른 바이러스량의 감소를 보였음.
Earnest, R. 2021	USA	Nasal swabs, anterior nares	Relevant	Between January 1, 2021 and August 1,	33,408 genomes	N1, N, E, ORF1a	Associated with the difference between Alpha and	델타 변이는 감염 속도가 빠르며, 전파력이 높고, 감염자들 또한 높은 바이러스량이 검출됨.

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
		swabs		2021			Delta variant's transmissibility	미국 New England 6개 주의 33,408 genome 대상으로 분석한 결과, 델타는 기존 알파 감염보다 6.2배 많은 viral RNA copy를 만들고 감염재생산지수 또한 63% - 167% 가량 높은 것으로 확인됨.
Chung, E. 2021	USA	Unsupervised home self-collected nasal swabs	Not relevant	Between March 23, 2020 and November 9, 2020	555 patients(123 children / 432 adults)	Orf1b	Associated with the difference between symptomatic and asymptomatic	아동과 성인에서 바이러스량과 증상 사이 연관관계는 어떻게 다른지 확인. 증상 유무와 관계 없이 아동과 성인 사이 바이러스량의 유의한 차이는 보이지 않았음. 증상이 있는 경우 더 높은 바이러스량을 보였음.
Kociolek, L. K. 2020	USA, Canada	Upper respiratory specimens(either NP or oropharyngeal [OP])	Not relevant	Between March, 2020 and July, 2020	817 patients (0 to 17 years) / 339 asymptomatic, 478 symptomatic	NA	Associated with the difference between symptomatic and asymptomatic in children	COVID-19 무증상 확진자 중 아동의 상부호흡기계의 바이러스량에 대해 알려져있는 바 없음. Ct value를 통해 확인한 바이러스량에서, 무증상인 경우 유증상 아동보다 더 낮은 바이러스량을 보였으나 진단 시점이 영향을 미쳤기에 주의해서 볼 필요가 있음.
Saththasivam, J. 2021	Qatar	Raw municipal wastewater	Not relevant	Between 21st June and 30th August 2020	Wastewater samples (n = 43)	N1	Predict future outbreak	카타르의 하수 기반 역학을 활용한 결과, 하수의 Ct value trend는 일일 확진자 증감과 같은 경향을 보였음.
Laresgoiti-Servitje, E. 2021	Mexico	Nasopharyngeal /oropharyngeal swabs	Not relevant	Between mid-April, 2020 and mid-September, 2020	1126 pregnant women (298 pregnant women with tested positive for Covid-19)	E, RdRp	Associated with presence of fibrinoid in pregnant women	임신한 여성에서 Ct value와 연관성 있는 지표들의 확인. 임신부에서 낮은 Ct value는 태반의 fibrinoid와 높은 연관성을 보이는 것으로 확인됨.

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
Weinbergerova, B. 2021	Czech	Nasopharyngeal swab sample	Not relevant	Between April 20, 2020 and September 2, 2020	105 cohort outpatients	E, RdRp	PCR-measured viral load kinetics	체코의 COVID-19 확진자들을 대상으로 covid-19의 자연사를 확인. 바이러스량은 증상 발현 기준 0-2일 내 제일 높은 것으로 확인됨.
Andriamandimby, S. F. 2021	Madagascar	Nasopharyngeal and oropharyngeal swabs	Not relevant	Between 18th March, 2020 and 30th September, 2020	34,563 RT-qPCR tests(20,326 discrete samples, 17,499 discrete patients)	E, N, ORF1a/b, or S	Predict future outbreak	마다가스카의 국립 열성 질환 연구소가 Ct value를 이용해 감염병 유행 양상을 예측함. Ct value가 낮을 때 유행병이 최고의 증가율을 보였으며, Ct value를 이용해 개인이 감염된 이후 경과한 시간에 대한 예측이 가능하다는 것을 확인함. 이를 기반으로, 인구집단의 Ct value 분포를 이용해 인구집단 전체의 역학 dynamic을 예측할 수 있다는 것을 밝힘.
Walker, A. S. 2021	UK	Nose and throat swabs and blood samples	Relevant	Between 26th April, 2020 and 13th March, 2021	440,479 participants	N, S, ORF1ab	Predict future outbreak	Ct value를 이용한 지역사회 유행병 감시체계의 가능성을 확인. 영국의 27,902 개 PCR 양성 샘플을 이용하여, Ct value 분포가 검사 양성률과 지니는 연관성을 확인함.
De Smet, D. 2021	Belgium	Nasopharyngeal swabs	Relevant	NA	study cohort (n = 397)	E, N, RdRp, S (S/RdRp)	Effective for variant diagnosis	B.1.351 변이에 대해 delayed N-gene amplification이 어떤 유용성을 가지고 있는지 확인. delayed N-gene amplification은 신속하고 정확한 B.1.351 스크리닝에 활용 가능하다는 것을 확인함.
Jmii, H. 2021	Tunisia	Wastewaters	Not relevant	Between 8th September, 2020 and 2nd October, 2020	31 raw wastewater samples	E, N, RdRp	Correlated with the number of daily cases	튀니지에서 하수를 이용한 COVID-19 모니터링 연구를 시행함. 두 번의 기간(2020년 9월-10월, 2021년 2월-4월)에 걸쳐 샘플이 수집되었으며, 하수에서 검출되는 viral RNA 양이 높을수록 COVID-19 확진자 또한 증가하는 양상을 확인함.
Anaclerio, F.	Italy	Nasopharyngeal	Relevant	Between April, 2020	150,000	ORF1ab, N,	Effective for variant	COVID-19 변이에 대한 추적과 대응은 중요한

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
2021		swabs		and April, 2021	nasopharyngeal swabs	S	diagnosis	관심사임. 변이를 추적하기에 제일 적합한 검사 방법에 대한 연구.
Bhargava, A. 2021	China	NA	Not relevant	NA	49 patients	ORF1ab, N	Interrelationships between ORF1ab and N	ORF1ab와 N 각각의 Ct value의 변이의 연관성과 환자의 변이를, 통계학적 방법으로 확인함. N RNA에 비해, ORF1ab RNA가 더 높은 환자 내부의 변이를 보임. ORF1ab RNA가 더 적합하게 활용될 수 있음.
Calistri, P. 2021	Italy	Nasopharyngeal /oropharyngeal swabs	Not relevant	Between 16th March, 2020 and 15th December, 2020	132,357 naso-pharyngeal/oropharyngeal swabs	N, S	Useful for monitoring the levels of viral load circulation in the population	이탈리아 Teramo 지역에서 실시한 데이터를 이용하여 Ct value의 분포를 확인함. 사회 전체에 대해 락다운이 시행되던 시기와, 그렇지 않은 시기의 Ct value 분포를 비교했을 때 중위값은 확연한 차이를 보임.
Hay, J. A. 2021	USA	Nasopharyngeal specimens	Relevant	Long-term care facilities data: Between 6th April, 2020 and 5th May, 2020 / Brigham and Women's Hospital data: Between 3rd April, 2020 and 10th November, 2020	NA	N1, N2, ORF1ab	Effective for estimating the epidemic growth rate and robust epidemic monitoring	확진자 수, 검사양성률, 사망자 수, 입원자 수에 기반한 유행병 양상 예측은 종종 편향된 결과를 제공할수도 있음. 반면, 무작위적으로 실행한 virologic survey는 이런 현행 감시체계들이 갖고 있는 단점을 보완하고 현재 감염병 유행 양상에 대한 정확한 정보를 제공할 수 있음. COVID-19의 viral load kinetics에 대해 미리 알려진 내용을 수리학적 모델에 대입하고 시뮬레이션해본 결과 Ct값의 중위값과 왜도를 통해 epidemic 양상 예측할 수 있다는 것을 보임.
Kennedy-Shaffer, L. 2021	UK	NA	Relevant	NA	Simulated samples	NA	Effective for estimation of the crucial measure of vaccine efficacy against transmission	COVID-19 백신의 감염과 중증화 진행에 대한 예방 효과에 더해, 감염 전파에 대한 효능 또한 점검해야 함. 이를 위해 바이러스량을 어떻게 활용할 수 있을지, 필요한 여러 가정과 그 가정이 위반되었을 때 대한 해결책을 제시함.
Valent, F.	Italy	Nasopharingeal	Not	Between March,	715 patients	NA	Associated with severity	이탈리아의 감염병 유행시기 별로 확진자들 중

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
2021		swabs	relevant	2020 and September, 2020				높은 바이러스량을 지닌 사람의 분율이 얼마나 되는지 확인. Ct value 20을 기준으로 하였을 때 높은 바이러스량 가진 분율은 2차 유행에서 1차 유행보다 2.9배 더 높은 것으로 확인됨.
Magleby, R. 2020	USA	Nasopharingeal swabs	Not relevant	Between 30th March, 2020 and 30th April, 2020	678 NP swab samples	ORF1ab, E	Associated with severity and death	COVID-19 확진자들을 Ct value <25, 25-30, >30 세 가지 기준으로 나누고, 각 집단에 속하는 환자들의 특성을 비교함. 높은 바이러스량(Ct value < 25)은 연령, 동반이환된 질환, 흡연, 최근 화학요법 치료와 연관성이 있었으며 기관삽관, 사망률의 위험이 높은 것으로 확인됨.
Boogaerts, H. L. F. 2021	Belgium	Nasopharyngeal swabs	Relevant	May, 2021	30 crew members	ORF1ab, N, S	Associated with severity and viral clearance with time	벨기에 안트베르펜의 선원들에서 발생한 집단감염을 대상으로 CT value와 viral culture를 실시함. CT value에 따른 이후 중증도(병원 입원)와 CT value와 viral culture 양성의 상관도 확인. CT value가 낮을수록 viral culture 양성이 오랜 기간 확인되고 병원 입원의 유효한 예측인자라는 것 또한 확인.
Yin, N. 2021	Belgium	Nasopharyngeal swabs	Relevant	Between April 1, 2020 and May 15, 2021	155,049 tests	RdRp, N	Predict future outbreak	벨기에 국가 수준에서, Ct value가 COVID-19 중환자 수를 예측할 수 있는지 확인. Ct value는 모든 연령군에서 유병률과 음의 상관성을 보임. 따라서 유행병 감시체계 지표로 활용할 수 있음.
Mack, C. D. 2021	USA	Nasopharyngeal swabs	Not relevant	Between 1st August, 2020 and 14th November, 2020		ORF1ab, E	Effective for early detection of infection	미국의 미식축구 연합회 대상으로 COVID-19 매일 모니터링을 실시한 뒤 결과를 확인. 조기 감염에서 높은 Ct value가 확인되었으며, 감염병의 조기 모니터링에 대해 PCR 검사가 유효하다는 것을 확인함.
Miller, E. H.	USA	Nasopharyngeal	Not	Between March 10,	5467	ORF1ab	Associated with mortality	COVID-19 바이러스량이 환자의 증상

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
2021		test	relevant	2020 and May 30, 2020	patients			지속기간과 어떤 연관성이 있는지 확인. Ct value가 낮을수록, 외래환자와 입원환자 모두에서 사망률이 높았으며 다른 통제변수들을 포함한 분석에서도 결과는 유의했음.
Bustos, P. 2021	Chile	Nasopharyngeal swab	Not relevant	Between March 2, 2020 and May 5, 2020	458 patients	ORF1ab	Help to monitor the SARS-CoV-2 infection in mild COVID-19 cases	Ct value를 통해 확인한 바이러스량과 질병의 중증도가 어떤 연관성이 있는지 파악하고자 함. 발열은 높은 바이러스량과 연관성을 보였으며, 바이러스량은 60세 미만 경증환자에서 증상 발현 후 1주 이내에 최정점을 기록함. 60세 이상 환자들은 증상 발현 후 2주까지도 최정점을 기록할 수 있음.
Griffin, J. B. 2021	USA	NA	Relevant	Between May 1, 2021 and July 25, 2021	9,651,332 Los Angeles County residents	ORF1ab, N	Correlated with the amount of viral nucleic acid present, but limitedly	Los Angeles County의 COVID-19 감염자 중 백신 접종과 미접종자, 중증 진행률에 대해 비교해 보았고 백신 접종자들에서 중증 진행도와 감염률이 현저하게 낮음을 확인함. Ct value는 백신 접종, 미접종 확진자 간 유의한 차이를 확인하지 못함.
Wollschläger, P. 2021	Germany	Nasopharyngeal swab	Relevant	NA	48 consecutive samples	S, RdRp, N	Effective for variant diagnosis	PCR 검사의 N gene dropout이나 S, RdRp의 Ct value와 비교했을 때 값의 변화가 B.1.1.7 변이 진단에 유용한지 확인. B.1.1.7 확진자 sample에서 Ct value는 S와 RdRp에 비해 6-10 높은 값을 보임.
Makov-Assif, M. 2021	Israel	Nasopharyngeal swabs	Not relevant	Between June 1st and December 21st, 2020	34,658 patients	ORF1a, 19N, N	Associated with severity	병원이 아닌 지역사회 맥락에서 Ct value와 질병 중증도의 연관성을 확인함. 낮은 Ct value는 증상, 입원, 중증화, 사망률 모두와

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
								연관성을 보임. 기존에 알려진 위험요인들을 보정했을 때에도 결과는 유의했으며, 연속변수로 분석하였을 때 양-반응 관계 또한 확인함.
Waudby-West, R. 2021	UK	Nasal and pharyngeal swab	Not relevant	All positive PCR test: between 12th March, 2020 and 1st May, 2020 / all deaths: between 12th March, 2020 and 20th May, 2020	1,337 patients	RdRp	Associated with mortality	Ct value를 >30, 20-30, <20 세 가지 분류(상,중,하)로 범주화하고 사망률과 연관성이 있는지 확인함. 낮은 Ct value는 높은 Ct value와 비교했을 때 높은 사망률을 보이는 것으로(HR 2.10) 확인됨.
Machado, A. 2021	Portugal	Nasopharyngeal swab	Not relevant	Between 17th March, 2020 and 1st June, 2020	197 patients	ORF1a/b, E	Associated with mortality	Ct value가 입원환자들의 사망률 증가와 연관이 있는지 확인. 입원환자들을 대상으로 사망률과 연관성이 있는 지표들을 확인하였고, E-gene Ct value, 연령, 입원 전 유증상으로 보낸 일수, LDH, SiO2/FiO2 비가 30일 사망률과 유의한 연관성을 보이는 것으로 확인됨. 바이러스량과 질병 중증도, ventilation, 재원기간은 유의한 연관성이 파악되지 않음.
Marks, M. 2021	Spain	Nasopharyngeal swabs	Not relevant	Between March 17 and April 28, 2020	314 patients	NA	Associated with the degree of risk of transmission and expected time to developing symptomatic COVID-19	COVID-19 감염 전파에 있어서 viral load가 어떤 영향을 미치는지 확실히 밝혀진 바 없음. 바이러스량과 연령, 성별, 증상 발생 이후 경과기간, 발열, 기침, 호흡곤란, 비염, 후각상실 등 특징과 연관성을 확인함. 또한 COVID-19 감염 cluster들을 대상으로 지표 환자가 바이러스량이 높을 때 2차 발병률이 높은 것을 확인함.
Kissler, S. M. 2021	USA	Anterior nares and	Not relevant	Between June 23, 2020 and Sep 7,	68 individuals	N1	Inform clinical measures and interventions	COVID-19 환자에서 시간에 따른 viral RNA shedding의 변화를 관찰함. 46명의 확진자를

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
		oropharyngeal swabs		2020				대상으로 시간에 따른 viral concentration 변화를 확인함. viral concentration은 평균 3.3일 이후 최정점에 도달하였으며, Ct value는 평균 22.3 값을 보임. viral clearance는 유증상자(10.9일)에서 무증상자(7.8일) 더 오래 걸렸고 첫 PCR 확진 2일 후 추가 PCR 검사는 확진자의 감염 단계에 대한 확실성을 향상시켰음.
Mohieldin Mahgoub Ibrahim, M. 2021	Italy	Respiratory samples(oro/nasopharyngeal swabs)	Not relevant	First period: between February, 2020 and May, 2020 / second period: between June, 2020 and August, 2020 / third period: between September, 2020 and February, 2021	254,744 samples(20,188 positive)	N1, N2	Associated with sex and age	이탈리아 북부 지역에서, 3개의 COVID-19 유행시기(2020년 2-5월, 6-8월, 9월 - 2021년 2월)에 따른 확진자들의 Ct value 분포와 검사 양성률을 비교함. 첫 번째 시기에서 Ct value가 제일 낮은 값을 보였으며(평균 25.64) 여성보다 남성에서, 고령일수록 더 낮은 Ct value를 보였음.
Knudtzen, FC. 2021	Denmark	Naso- and/or oropharyngeal swab, or sputum	Not relevant	March 9 - March 17 2020	169	NA	Predict disease severity	입원환자와 외래환자들을 대상으로 Ct value가 중증도와 연관성이 있는지 확인함. 입원 환자에서 Ct value가 낮을수록 질병의 중증도가 심해지는 결과를 확인함(aOR 0.89, 95% CI 0.81-0.98).
Garg, K. 2021	India	NA	Not relevant	NA	NA	N2, E	Predict severity and outcomes in patients	Ct value를 high (<25), medium (25.1-32) and low (>32) 세 분류로 나누고 질병 사망률과 중증도의 연관성을 확인함. Ct value는 중증도와 사망률 모두에서 역의 연관성을 보임.
Pasculli, A. 2021	Italy	Nasopharyngeal swab	Not relevant	March 4 to May 31, 2020	283	E, S	Predict prognosis in COVID-19 patients	Ct value와 CT 영상 촬영물을 함께 활용할 경우 COVID-19 환자에서 예후를 예측하는 데

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
								도움이 될 수 있음.
Gentilini, F. 2021	Italy	Nasopharyngeal /oropharyngeal swab	Not relevant	NA	51	NA	To exploit the features of digital PCR	6208개의 COVID-19 진단 샘플을 이용해 Ct value를 바이러스량(copies/μL)으로 환산하는 식을 모델링함.
Abeysuriya, V. 2021	Sri Lanka	Nasopharyngeal swab	Not relevant	May 2020 - April 2021	118	NA	To predict respiratory compromise	Ct value, 림프구 수(AL), 백혈구:림프비비(NLR)를 이용해 저산소증을 예측할 수 있는지 확인함. Ct value가 19.9 미만인 경우 호흡부전이 생길 확률이 높은 것으로 확인됨.
Stankiewicz, H. 2021	USA	NA	Relevant	March 31 to August 21, 2020	829	NA	To characterize The SARS-CoV-2 viral trajectory	감염 초기의 COVID-19 viral load를 확인하기 위해 종단적으로 관찰한 확진자들의 Ct value를 활용함.
Skarzynski, M. 2021	USA	NA	Relevant	16-Apr-21	NA	N1	To rapidly identify which genetic variants result in increased virulence or transmission	Ct value와 genome-based predictive algorithm을 이용해 716개 viral genome을 대상으로 임상 중증도를 예측하는 모형을 개발함. Ct value의 중위값이 16.6 이하일 경우, 21.4 이상인 경우보다 임상적 예후가 악화될 확률이 현저히 높았음.
Kurzeder, L. 2022	Germany	NA	Relevant	February 1st to May 31st 2021	NA	E	To develop a simple score for the outcomes	입원 당시 Ct value와 eGFR, 산소 포화도, 연령은 COVID-19 환자의 사망률을 예측하는데 유의한 결과를 보였음.
Deng, Y. 2022	Hong Kong	Wastewater	Relevant	2021 June	NA	N1, E	To initiate public health action to thwart a looming Delta variant	30,000 명이 거주하는 지역의 하수 샘플에서 시행한 routine sewage surveillance를 통해, 델타 변이 확진자의 발생을 미리 예측할 수 있었음.
Valley-Omar, Z.	South Africa	NA	Relevant	2021 Jan-Aug	2297	RdRp	Predict Delta variant prevalence	delta 변이 감염자가 CT value가 RT-PCR amplification efficiency를 감소시킨다는 것에

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
2022								착안하여 델타 변이 유병률을 예측하고자 함.
Trobajo-Sanmartín, C. 2022	Spain	NA	Relevant	NA	NA	S, RdRp, N	Evaluate the analytical performance of assays	COVID-19 variant 별로 N, RdRp, S gene을 target으로 한 Ct value의 변이가 생기는 것을 이용하여 B.1.1.7과 B.1.177 변이에 대한 예측지표로 Ct value 활용 가능성을 고려함.
Makulo JR. 2022	DR Congo	NA	Not relevant	June 1st, 2020 - August 02, 2020	242	N2, E	Evaluate the concordance between RT-PCR and serological tests (IgM/IgG),	Ct value의 사망률에 대한 영향을 확인하였으나, CtE와 CtN2 값 모두 사망률에 유의한 영향을 미치지 않는 것으로 확인함.
Alizon, S. 2022	France	NA	Not relevant	21 January - 30 November 2020	793,479	NA	Use Ct value to predict epidemic trends	프랑스의 Ct value 793,479 건을 활용하여 인구집단 수준에서 Ct value의 전체 분포가 epidemic trend를 예측할 수 있는지 확인. 유행 양상은 Rt 값을 통해 예측하였으며, Rt 값을 예측하는 ARIMA 모델을 개발한 뒤 Ct value 분포에 대한 정보를 포함시킬 경우 예측력이 향상되는 것을 확인함.
Lin, Y. 2022	Hong Kong	NA	Not relevant	NA	NA	NA	Use Ct value to predict epidemic trends	기존 Rt 값에 기반한 유행 규모 예측은 1주 간의 lag이 생긴다는 단점이 있음. 실시간 예측을 활성화하기 위하여 확진자들의 Ct value를 이용한 예측을 시도함. Ct value의 분포와 epidemic 진행이 유의한 상관성을 지님을 확인하였으며, Ct value를 활용해 유행 규모를 예측할 경우 그 예측력이 향상되는 것을 확인함.
Tharak, A. 2022	India	Wastewater	Not relevant	July 2020 to August 2021	NA		To predict the early surge of pandemic	하수의 viral load의 시계열적 변화를 통해 지역사회의 COVID-19 유행 양상과 규모를 예측함.
Yuniarti, L.	Indonesia	NA	Not	December 2020 to	286	N,E	Analyze the association	COVID-19 환자의 중증도를 예측하는 여러

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
2022			relevant	March 2021			between the severity	인자들에 대한 연구. 연령, 성별, 질병이환 등 요소와 더불어 낮은 Ct value가 환자들의 질병 중증도와 상관성이 있다는 것을 확인함.
Amereh, F. 2022	Iran	Sewage	Not relevant	late September 2020 until early April 2021	NA	N, ORF1ab	Regular monitoring	하수 sample을 대상으로 시행한 RT-qPCR에서 계산한 바이러스량이 COVID-19 유행과 양의 상관성을 보였으며, 실제 인구집단에서 감염이 전파되기보다 1-2일 앞서 유행을 예측함. 따라서 실제 공중보건학적 정책 결정에 활용할 수 있을 것으로 보임.
Trunfio, M. 2022	Italy	Nasopharyngeal swab	Not relevant	2020 Mar	NA	RdRp	Identifying determinants of the COVID-19 transmission in settings of contagion	유증상 성인 COVID-19 환자의 가구 내 감염 추가 전파와 관련된 요인에 대한 연구. Ct value를 범주화하여 감염 추가 전파와 연관성을 확인하였으며 Ct 값이 높을수록 추가 감염 전파가 될 가능성이 낮다는 것을 확인함.
Zhu, Y. 2022	Japan	Wastewater	Not relevant	August 2020 to February 2021	NA	NA	Time series of SARS-CoV-2 RNA occurrence in the wastewater	2020년 8월부터 2021년 2월까지 360,000 명이 거주하는 도심 지역사회에서 하수의 COVID-19 viral load를 확인함. 유행이 있을 때 하수의 바이러스량 또한 증가하는 것을 확인함.
Proverbio, D. 2022	Luxembourg	Wastewater	Not relevant	NA	NA	NA	Predict new cases	하수의 viral load와 신규 확진자 수를 함께 활용하여 역학적 예측 모델을 형성하고 이를 기반으로 COVID-19 신규 확진자 수 예측이 가능하다는 것을 보임.
Barajas-Carrillo, V. 2022	Mexico	Nasopharyngeal /oropharyngeal swab	Not relevant	1 August 2020; 21 January 2021, 19 August 2021, and 17 January 2022, respectively	158, 254, 143, and 279 cases in each	NA	Understand the distribution dynamics within a defined geographic area	지역의 COVID-19 전파 파악을 위해 GIS를 이용한 분석을 시행. 개인의 Ct value를 범주화하고, 이를 기반으로 공간분석을 시행하여 감염 전파가 위험한 곳을 확인하는 방법을 통해 정책 의사결정에 도움이 될 수

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
								있음.
Shoaib, N. 2022	Pakistan	NA	Not relevant	May 1, 2020 - November 30, 2020	6331	Orf1ab, N, and RdRp	To find indicators associated with Ct value	COVID-19 환자들을 대상으로 Ct value와 연관성 있는 지표들을 확인함. 6331명을 대상으로 한 연구에서 연령, 성별, 임상 기록과 Ct value는 유의한 연관성을 보이지 않았으며 ORF1ab, N gene Ct value만이 검사 당시 증상과 약한 연관성을 보임.
Dergaa, I. 2022	Qatar	Nasopharyngeal /oropharyngeal swab	Not relevant	2020 July	481	NA	Investigating Association with clinical factor	카타르의 COVID-19 유증상 확진자와 무증상 확진자들을 대상으로 Ct value를 통해 파악한 viral load와 연관성 있는 지표들을 확인. 무증상자에 비해 유증상자가 Ct value, 산소포화도가 낮았으며 더 높은 체온을 보임. 높은 바이러스량은 높은 심박수, 체온, 낮은 산소포화도와 연관성이 확인되었으나 연령과 연관성은 확인되지 않음.
Pérez-García, F. 2022	Spain	Nasopharyngeal swab	Not relevant	9 November 2020 - 8 March 2021	255	NA	Association with severity	상부 호흡기의 점액 면역 반응과 COVID-19 진행 및 예후의 연관성을 확인함. 상부 호흡기에서 높은 바이러스량과 낮은 CCL5 expression은 질병의 중증도와 유의한 연관성을 보임.
Vallejo, J. 2022	Spain	Wastewater	Not relevant	NA	NA	NA	Association with severity	하수의 RNA viral load를 정량화하여 지역사회의 COVID-19 outbreak을 감시하는데 활용하고자 함. 지역의 역학 데이터와 하수 데이터를 함께 활용할 경우 유증상자와 무증상자를 모두 포함한 지역의 COVID-19 유행 규모를 예측할 수 있다는 것을 확인함.
Ingberg, E. 2022	Sweden	NA	Not relevant	1 March - 31 August 2020	286	N, E, RdRp	Predict outcome	COVID-19 환자에서 흉부 CT를 통해 확인한 폐 침엽 정도와 Ct value의 연관성을 확인하고, 두 가지 지표를 결합하여 질병의 중증 진행률을 예측함. Ct value와 흉부 CT 모두 중환자실

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
								입원과 사망에 대한 유의한 연관성을 보임.
Bolze, A. 2022	USA	NA	Relevant	Jan – Sep 2021	583,984	NA	To find whether Delta displace variant Alpha, Gamma, Iota, and Mu variants	2021년 3월부터 미국 전역의 우세종이 된 델타 변이의 Ct value를 분석함으로써 다른 변이와 viral load의 차이를 확인. 델타 변이는 알파 변이보다 1.7배 높은 바이러스량을 보였으며, 우세종이 되는데 높은 바이러스량을 형성하는 특징이 작용하는 것으로 고려됨.
Hawken, S. 2022	USA	NA	Not relevant	17 March 2020 - 15 October 2020	264	NA	Identify characteristics of longitudinal positivity prior to vaccination and VOC circulation	백신 접종과 VOC 출현 이전 장기적으로 COVID-19 바이러스가 배출되는 환자들을 대상으로 Ct value를 확인. 임상 증상 맥락 없이 Ct value를 중증도, 감염성을 반영하는 지표로 해석하는 것은 주의가 필요함.
Hoar, C. 2022	USA	Wastewater	Not relevant	November 8, 2020 to April 11, 2021	NA	NA	Predict cases for the populations by wastewater	2020년 11월부터 2021년 4월까지 NYC의 하수에서 파악한 viral load를 COVID-19 신규 확진자 수 예측에 활용할 수 있는지 파악함. 신규 확진자 수와 하수의 viral load는 유의한 상관성을 보였으며 회귀분석에서도 유의한 결과를 확인함.
Morris, C. 2022	USA	NA	Relevant	From March 11th 2020 to September 23rd 2021	1529	NA	Identify patients with long viral shedding	대규모 COVID-19 molecular testing과 whole genome sequencing은 실시간 genomic surveillance에 활용할 수 있음. prolonged RNA shedding을 보이는 환자들을 대상으로 Ct value와 genome sequencing을 시행하였고 백신 미접종자와 면역저하자에서 viral shedding이 더 오래 지속되는 것을 확인함.
Wang, Y. 2022	USA	Wastewater	Not relevant	July 13th, 2020 - March 14th, 2021	NA	NA	Predict COVID-19 cases	미국 Emory 대학의 병원, 격리 건물, 기숙사 건물에서 시행한 wastewater surveillance의 결과를 발표. 기숙사 건물의 하수에서 viral load가 상승할 경우 1-2주 이후 COVID-19 확진자 수가 급증하는 것을 확인함. 따라서

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
								기관에서 wastewater-based surveillance의 가능성을 확인함.
Wünsch, K. 2022	Germany	Nasopharyngeal swabs	Not relevant	December 2020 - April 2021	79	(1) S, E (2) RdRp, N	In-hospital management of high-risk COVID-19 patients	고위험(60세 이상, 면역저하 또는 당뇨병 환자) COVID-19 환자에 대한 병원 내 관리방안으로 양성 SARS-CoV-2 RNA 검출에 관계없이 대부분의 샘플에서 사이클 임계값(ct)인 33을 넘어 배양할 수 있는 바이러스는 없었으므로, 노인과 당뇨병 환자는 ct 값 35 이상에서는 안전하게 격리될 수 있음.
Wu, J. 2022	China	Throat swab	NA	NA	245 patients from 60 cluster infections	ORF1ab, N	Important factor for disease severity and potential transmission capability of a patient	바이러스 부하(viral load)는 환자의 질병 심각도와 잠재적 전염 능력에 중요한 요인으로, N 유전자의 Ct 값은 임상 범주마다 유의하게 달랐음(특히 중증 환자의 경우 바이러스 양이 유의하게 더 높았음). 따라서 바이러스 부하가 더 높은 중증 환자는 그에 상응하여 더 높은 전파 잠재력을 가질 수 있음.
Souverein, D. 2022	Netherlands	Combined nasopharyngeal (NP) and oropharyngeal (OP) swabs	Not relevant	Between 17 March 2020 and 31 March 2021	20,207 individuals tested positive for SARS-CoV-2	E	Associated with an increased risk of hospital admission, ICU admission, and in-hospital mortality	초기 SARS-CoV-2 바이러스 부하가 높을수록 입원, ICU 입원, 병원 내 사망 위험이 증가함.
Singanayagam, A. 2022	UK	Self-collected upper respiratory tract (URT) swab	Relevant	Between September 13, 2020, and September 15, 2021	621 participants	ORF1ab, E	Fully vaccinated individuals with breakthrough infections can efficiently transmit infection in household settings, including to fully vaccinated contacts.	예방접종을 완료한 사람들의 돌파감염 사례는 예방접종 미접종 사례와 유사하게 최대 바이러스량을 가지며, 예방접종을 완료한 사람들 간의 접촉을 포함한 가정 환경에서도 효율적으로 감염을 전염시킬 수 있음.
Roversi, M. 2022	Italy	Nasopharyngeal swab (NPS)	Not relevant	Between May 2020 and April 2021	570 patients	(1) E, N (2) S, ORF1ab	Should not be used as a measure of clinical severity of COVID-19 in the pediatric	감염된 아동의 바이러스 양을 가장 잘 예측하는 임상 변수는 COVID-19 감염자에 대한 이전의 노출(exposure) 이력 및 발열 및 호흡기

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
						(3) E, RdRp/S, N	population	증상(높은 바이러스 양)임. SARS-CoV-2 바이러스 부하는 소아 인구에서 COVID-19의 임상적 심각도를 측정하는 척도로 사용해서는 안 되지만, 바이러스 부하의 감소는 연상의 아이들(older children)에서 무증상 COVID-19와 관련이 있는 것으로 보임.
Phillips, M. C. 2022	USA	Nasopharyngeal swab (NPS)	Relevant	Between October 1, 2020 and March 31, 2021, and between May 4, 2021-July 31, 2021.	2,114 SARS-CoV-2-positive samples	E, N2	To be a leading indicator for predicting future COVID-19 cases	COVID-19의 감염 증가는 일시적으로 더 높은 바이러스 부하와 관련이 있음. Ct 값 감소는 향후 COVID-19 사례를 예측하는 선행 지표로 나타나 병원 수준의 surge planning을 개선할 수 있음.
Nelson, J. R. 2022	USA	Wastewater	Not relevant	From May 2020 through January 2021	NA	N	Some locations wastewater loading values might precede case reports by 7–10 days	RT-qPCR을 통해 측정된 SARS-CoV-2 폐수 부하 수준과 함께 사용할 때 COVID-19 사례를 더 작은 분석 단위로 세분화하고 배분하는 것이 지역사회 확산에 관한 더 세밀한 정보를 어떻게 드러낼 수 있는지를 보여줌. 집계된 ZIP 코드 데이터는 1-2일 전에 사례 보고의 전조를 제공하지만, ZIP 코드 수준 분석은 일부 지역의 폐수 적재 값이 사례 보고보다 7-10일 앞서 있을 수 있음을 시사함.
Khalil, A. 2022	Lebanon	NA	Not relevant	Between 1 March 2020, and 31 March 2021	23,185 patients	NA	To now-cast the number of nationwide positive COVID-19 cases in a specific region	대표적인 횡단면 환자 코호트의 혈청학적 분석을 발생률 및 감염 규모와 같은 역학 지표와 결합하여 특정 지역의 전국적인 양성 COVID-19 사례 수를 예측함. 관찰된 바이러스 부하 측정을 활용하는 강력한 프레임워크를 사용하여 향후 7일 기간 동안 새로운 COVID-19 사례의 시계열 나우캐스팅을 함.
Karron, R. A. 2022	USA	Self-collected nasal swabs	Relevant	Between November 24, 2020, and October 15, 2021	690 participants	NA	The presence and number of symptoms did not correlate with viral load	COVID-19 감염은 0세에서 4세 사이의 어린이들 사이에서 종종 무증상이었음. 증상들의 존재와 수는 바이러스 부하와 상관관계가 없었으므로, 증상 선별은 어린

Authors	Country	Specimen Type	Variant Relevance	Length	Population	CT value target gene	Main outcome with CT value	요약
								아이들을 포함하는 발병을 예방하기에 불충분할 수 있음.
Hall, S. M. 2022	USA	Self-administered anterior nares swabs	Not relevant	Between August 7, 2020 and March 18, 2021	1,633 positive cases	N1, N2, Rnase P	Symptomatic individuals have higher viral loads than those who are pre- or asymptomatic	평균적으로, 증상이 있는 사람은 증상이 없거나 증상이 없는 사람보다 바이러스 부하가 높다는 주장을 뒷받침함.
de Sousa, A. R. V. 2022	Brazil	Influent(raw sewage) and final effluent (treated sewage)	Not relevant	From January to August 2021	Over 700,000 people	N1, N2	Correlation between the clinical data and the viral load detected	임상 데이터와 검출된 바이러스 부하 간의 상관관계를 관찰하고, 폐수 시료에 존재하는 바이러스 부하 검출을 기반으로 감염자 수를 예측하는 사용자 친화적인 웹 어플리케이션을 개발하여 진행 중인 outbreak을 모니터링하기 위한 공공 정책 전략으로 적용할 수 있음.
Coyle, P. V. 2022	Qatar	Combined nasopharyngeal/ oropharyngeal swabs (NPS/OPS)	Not relevant	NA	173,557 individual Ct values from 148,066 patients	(1) ORF1ab, E (2) E, N (3) N, S, ORF1ab	Improved understanding of SARS-CoV-2 results and impacted on patient and public health management criteria	일상적인 보고에서 Ct 값과 해석적 코멘트를 사용하면 SARS-CoV-2 결과에 대한 이해가 개선되었고 환자와 공중 보건 관리 기준에 영향을 미쳤음.
Chassalevris, T. 2022	Greece	Wastewater	Relevant	From November 11, 2021, to January 9, 2022	700,000 people	N2	Developing PCR-based WBE method to detect and quantify the Omicron VOC	Omicron VOC를 검출하고 정량화하기 위한 PCR 기반 WBE 방법을 개발함. Omicron은 임상 샘플에서 인식되기 훨씬 전에 폐수에서 검출되었음. 이 방법은 고도로 확산되는 VOCs(변종코로나19바이러스)에 대한 조기 경고에 적용될 수 있음.
Barry, M. 2022	Kingdom of Saudi Arabia	Nasopharyngeal (NP) and endotracheal (ET) samples	Not relevant	Between March 2 and September 1, 2020	728 patients	E	Predictors of IMV and mortality, and this effect increases when combined with age	입원한 COVID-19 환자들의 SARS-CoV-2 RT-PCR Ct 값은 IMV(invasive mechanical ventilation)와 사망률의 예측 변수로 사용될 수 있으며, 이러한 효과는 연령과 결합할 때 증가함. 임상의는 RT-PCR Ct 값으로 나이 든 환자를 위험 계층화를 위해 분류하고 IMV를 할당할 수 있음.

Appendix B. Time series distribution graph of Ct value RdRp by diagnosis variables

Figure B.1. Time series distribution graph of Ct value RdRp by diagnosis agency type (by date)

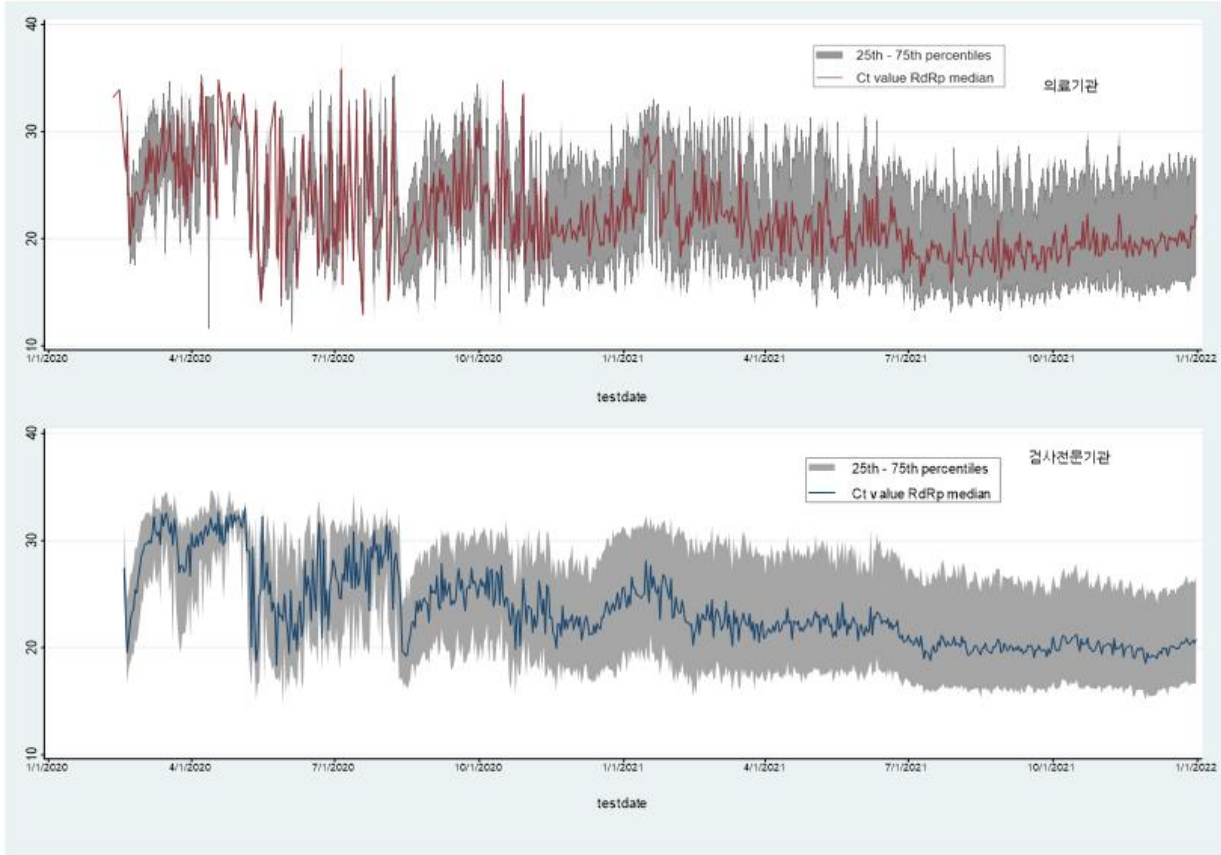


Figure B.2. Time series distribution graph of Ct value RdRp by agency Located regions (by date)

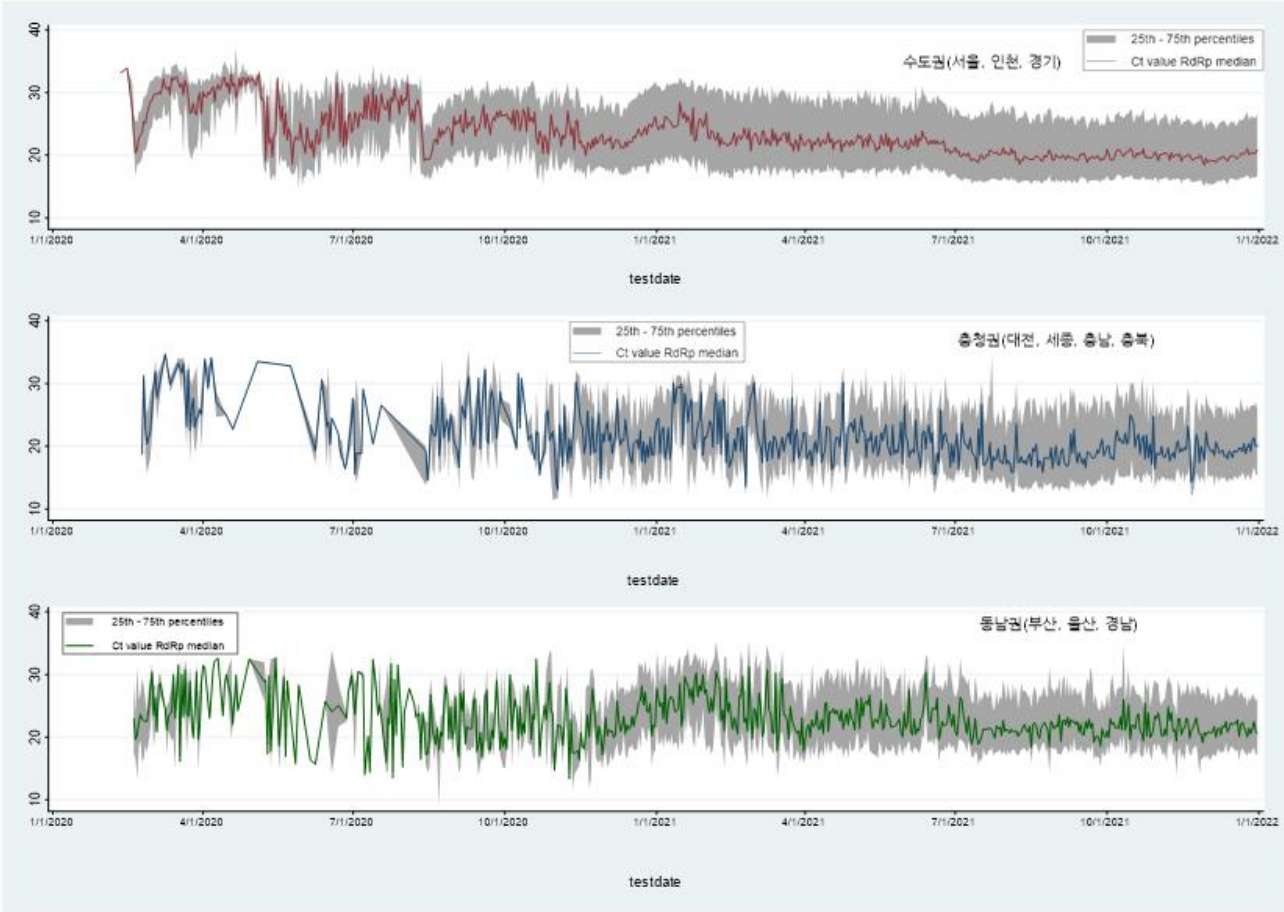


Figure B.3. Time series distribution graph of Ct value RdRp by age group (by date)

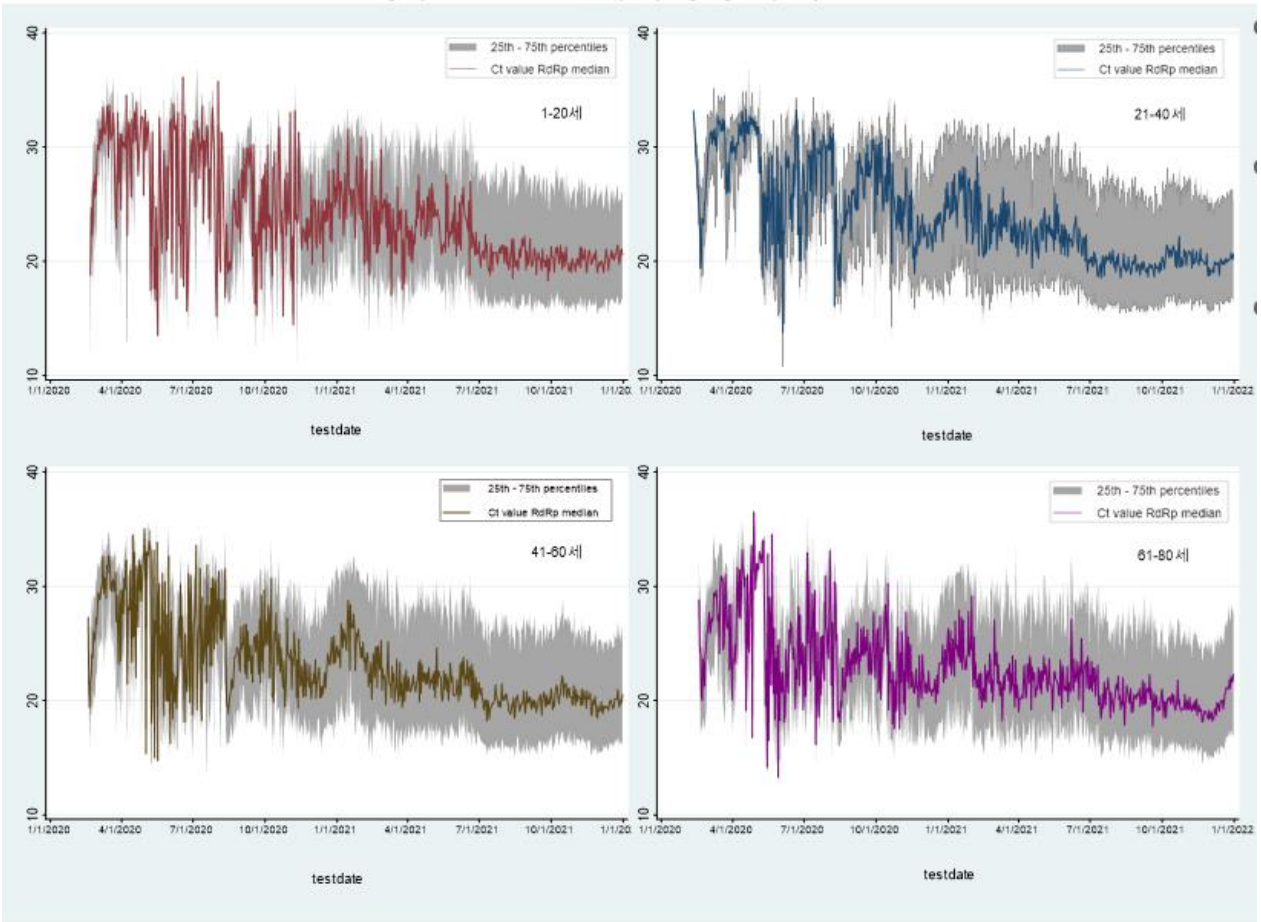


Figure B.4. Time series distribution graph of Ct value RdRp by sex (by date)

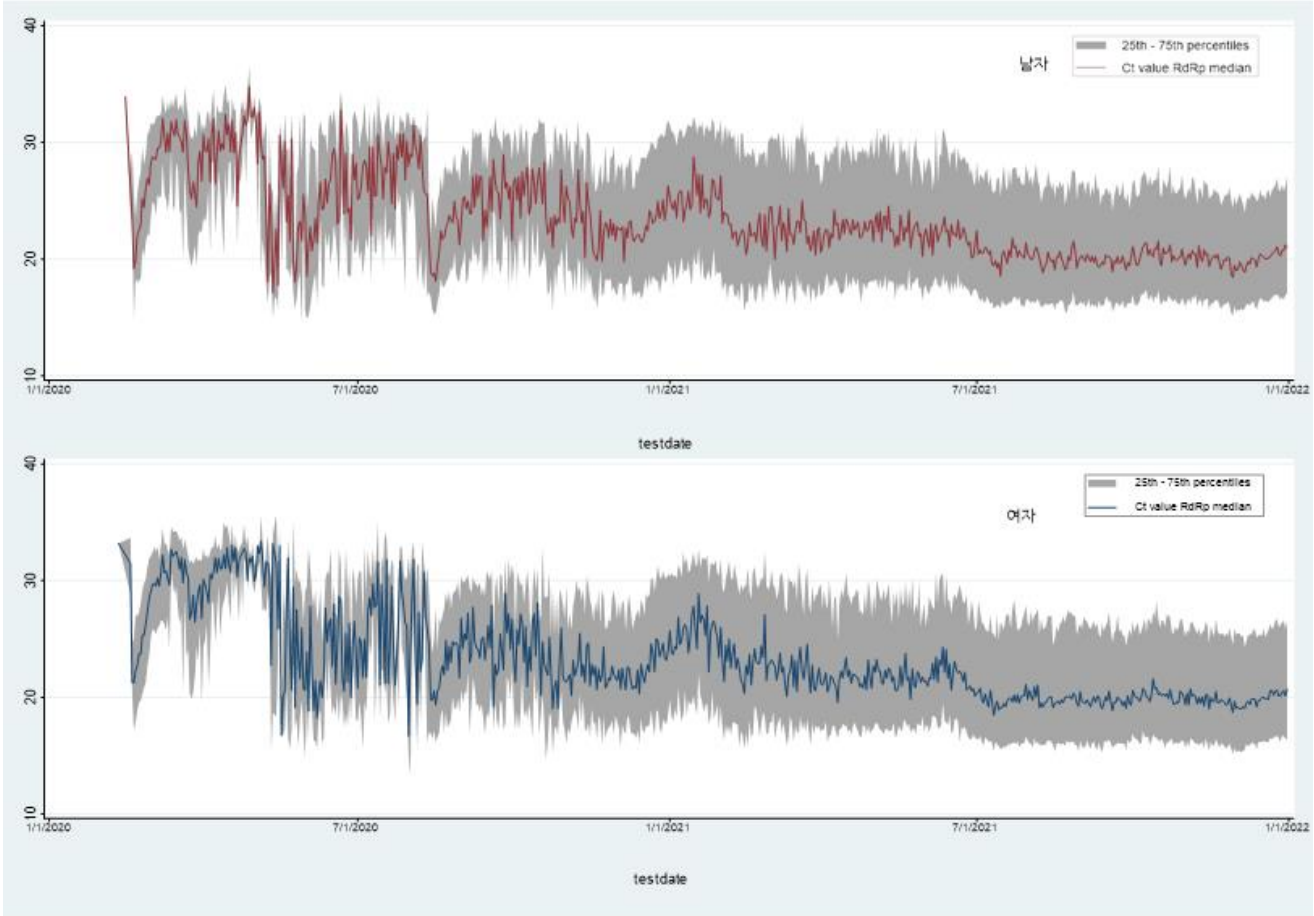
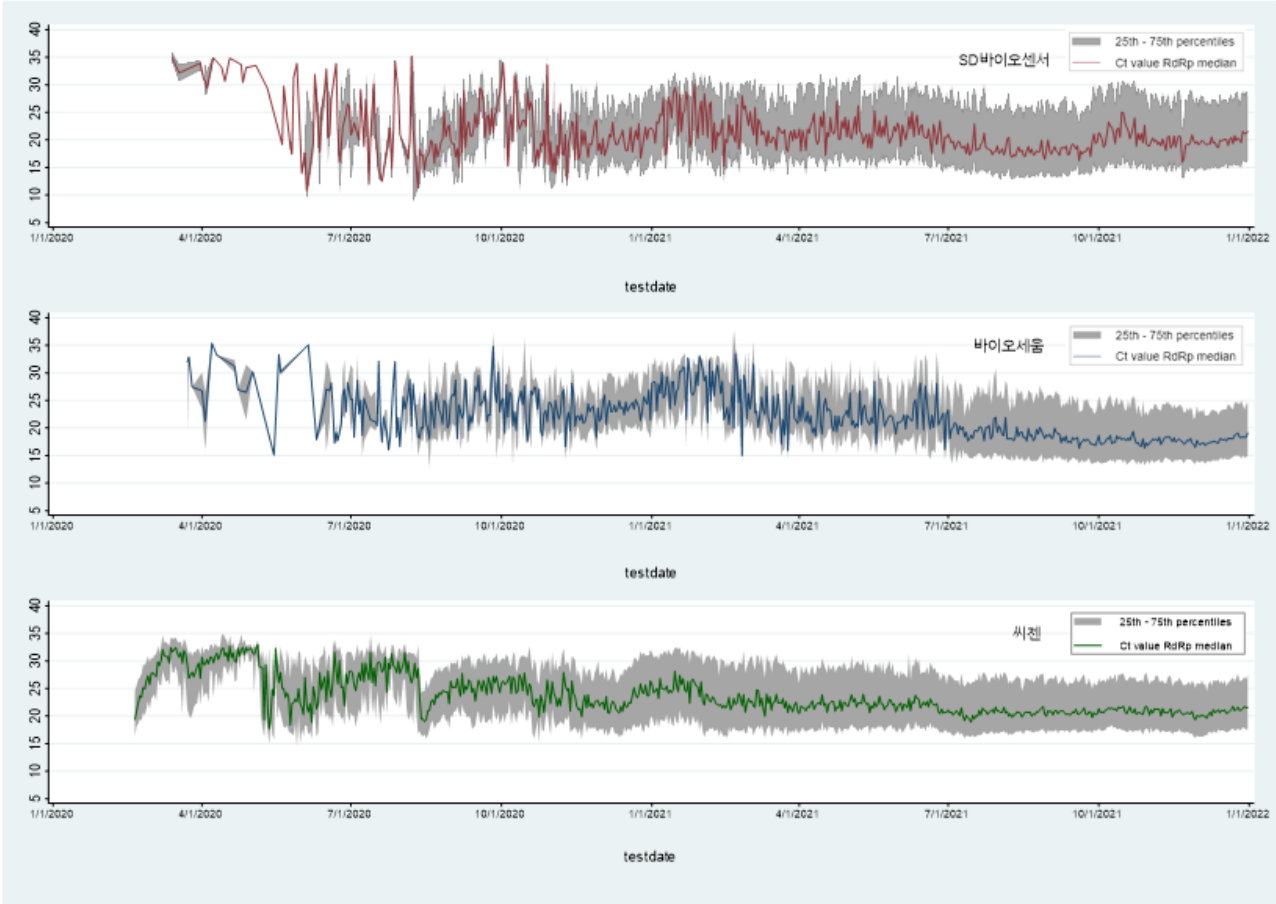


Figure B.5. Time series distribution graph of Ct value RdRp by assays (by date)



Appendix C. Unit root test results (Augmented Dickey-Fuller test)

Table C.1. Unit root test results of Ct value RdRp and number of newly confirmed cases

```
Augmented Dickey-Fuller test for unit root

Variable: ctRdRpmed                     Number of obs   = 681
                                        Number of lags  =   1

H0: Random walk without drift, d = 0

                                   Dickey-Fuller
                  Test      ──────── critical value ────────
               statistic          1%           5%          10%
──────────────────────────────────────────────────────────────
 Z(t)             -3.859      -3.430       -2.860       -2.570
──────────────────────────────────────────────────────────────
MacKinnon approximate p-value for Z(t) = 0.0024.
```

```
Augmented Dickey-Fuller test for unit root

Variable: newlyconfirm                  Number of obs   = 681
                                        Number of lags  =   1

H0: Random walk without drift, d = 0

                                   Dickey-Fuller
                  Test      ──────── critical value ────────
               statistic          1%           5%          10%
──────────────────────────────────────────────────────────────
 Z(t)             -1.805      -3.430       -2.860       -2.570
──────────────────────────────────────────────────────────────
MacKinnon approximate p-value for Z(t) = 0.3778.
```

Table C.2. Unit root test results of logged number of newly confirmed cases

```
Augmented Dickey-Fuller test for unit root

Variable: log_new                        Number of obs  = 681
                                         Number of lags =   1

H0: Random walk without drift, d = 0

                                     Dickey-Fuller
                  Test       ──────── critical value ────────
              statistic           1%           5%          10%
──────────────────────────────────────────────────────────────
 Z(t)            -2.158       -3.430       -2.860       -2.570
──────────────────────────────────────────────────────────────
MacKinnon approximate p-value for Z(t) = 0.2219.
```

Table C.3. Unit root test results of first difference of logged number of newly confirmed cases

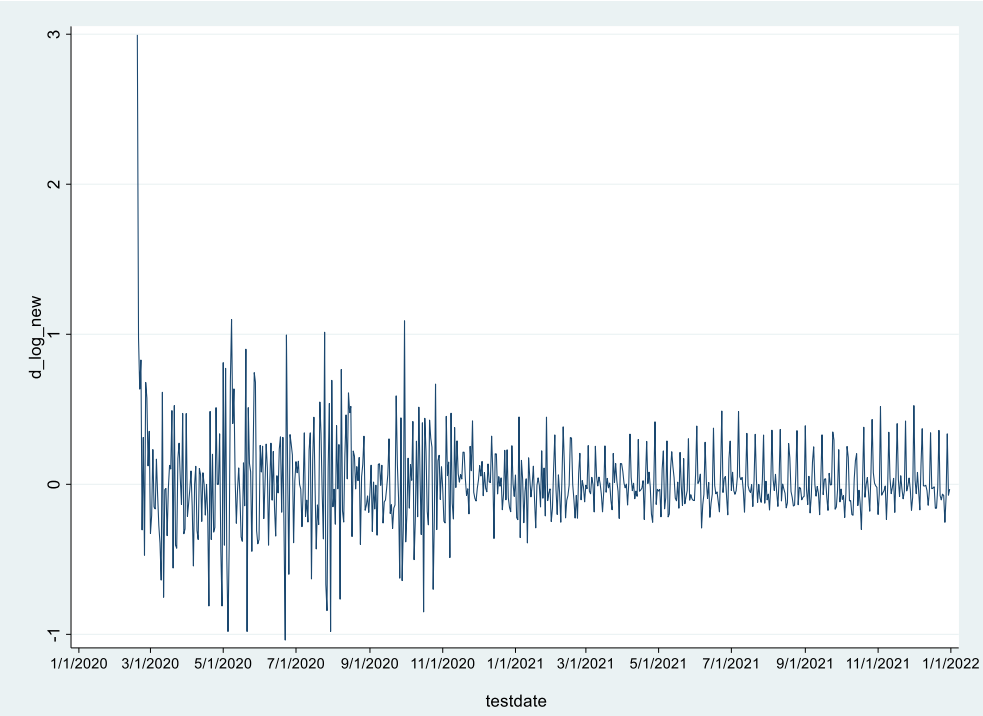
```
Augmented Dickey-Fuller test for unit root

Variable: D.log_new                      Number of obs  = 680
                                         Number of lags =   1

H0: Random walk without drift, d = 0

                                     Dickey-Fuller
                  Test       ──────── critical value ────────
              statistic           1%           5%          10%
──────────────────────────────────────────────────────────────
 Z(t)           -23.456       -3.430       -2.860       -2.570
──────────────────────────────────────────────────────────────
MacKinnon approximate p-value for Z(t) = 0.0000.
```

Figure C.3. Graph of the first difference of logged number of newly confirmed cases (by date)



Appendix D. Autocorrelation test results in the errors in a regression model (Breusch–Godfrey test)

Table D.1 Autocorrelation test results (by date, all periods)

Breusch-Godfrey LM test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
1	38.540	1	0.0000

H0: no serial correlation

Table D.2. Breusch–Godfrey test (by date, 1st COVID-19 periods)

Breusch-Godfrey LM test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
1	16.129	1	0.0001

H0: no serial correlation

Table D.3. Breusch–Godfrey test (by date, 2nd COVID-19 periods)

Breusch-Godfrey LM test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
1	10.343	1	0.0013

H0: no serial correlation

Table D.4. Breusch–Godfrey test (by date, 3rd COVID-19 periods)

Breusch-Godfrey LM test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
1	2.156	1	0.1420

H0: no serial correlation

Table D.5. Breusch–Godfrey test (by date, 4th COVID-19 periods)

Breusch-Godfrey LM test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
1	1.978	1	0.1596

H0: no serial correlation

Abstract

Epidemiological Application of Cycle Threshold Values Under The COVID-19 Pandemic in Korea

JEEWOUN KIM
Departments of Spatiotemporal Epidemiology
The Graduate School of Public Health
Seoul National University

Backgrounds Cycle Threshold values [Ct values] are the measured values, the determinants of positivity in a COVID-19 diagnosis. Studies on the usefulness of Ct values for forecasting the number of COVID-19 newly confirmed are being actively conducted worldwide because of their features that are inversely proportional to the viral load, and their association with infectivity/transmission (et al., 2021; Rodríguez et al., 2021). In contrast, Ct values have been used only in the diagnostic field in Korea. Forecasting the number of newly confirmed cases and public health strategies for responding to COVID-19 has relied on the time-varying reproduction number [Rt] from the beginning of the COVID-19 outbreak. Rt indicates how many additional confirmed cases one confirmed patient has created. However, Rt has some limitations. It is challenging to reduce Rt estimation to less than a week (Lin et al., 2022), and its definition and calculation formula require identifying the accurate infection route (Yoo, 2021; Jung, 2020). To improve the forecasting model of COVID-19 newly confirmed in Korea, this study conducted the forecasting with Ct values to supplement the limitations of Rt. As a result of a

systematic literature review, there were some differences between the results of static estimation and dynamic estimation. The research objectives were as follows. First, Rt and Ct values were compared through the time series distribution by COVID-19 pandemic periods and visually verified whether Ct values could forecast the actual new confirmation. Next, time series analysis-based forecasting of the COVID-19 confirmation pattern using the Ct values by COVID-19 pandemic periods was conducted to determine the time difference of several days when predictive power is the highest. This study conducted forecasting by using both static and dynamic estimation methods for comparison of each result.

Methods The study period included 489,133 specimens collected for SARS-CoV-2 diagnostic testing about two years from February 7, 2020, to December 31, 2021. Following domestic RT-PCR protocol, this study used Ct values RdRp as representative Ct values. Based on the Korea Disease Control and Prevention Agency [KDCA] publication, the periods of the COVID-19 pandemic in Korea were classified into four periods.

This study presented the descriptive statistics table and graph for each variable of the diagnostic test results to understand the overall diagnostic status of the samples included in the analysis. For visual verification before forecasting, the distribution pattern of the Ct values at the population group level by the COVID-19 pandemic periods is expressed as a time series graph along with the Rt and the number of new confirmed cases, respectively. After verifying the limitations of Rt and the utility of Ct values (as a forecasting index for new confirmation) through the two distribution graphs, a statistical analysis was conducted to confirm the actual predictive power of Ct values. To ascertain the time difference when Ct values show the highest predictive power to newly confirmed cases, this study conducted both static and dynamic estimation methods with time lags. A simple linear regression model was used as a static estimation method, and the Distributed Lag Model [DLM] was used as a dynamic estimation method. An alpha of 0.05 was used for all tests, and all graphs and statistical analyses were performed with STATA (version 17).

Results As a result of the descriptive statistics, most of the diagnostic tests were conducted at examination institutions, and more than half of the confirmed patients were in age from 20s to 50s. More than half of the diagnostic kit products

used were Seegene. In almost all tests, the same gene amplification equipment was used, and more than half of the samples used were throat swabs (NPS/OPS). Ct values E, RdRp, and N all showed similar distribution forms, and values were mainly distributed between 10-20.

In the two Ct values distribution pattern graphs, it was visually ascertained that the changes in Rt, compared to Ct values, did not sufficiently explain the third and fourth periods when the number of new confirmed cases increased rapidly. Also, it was ascertained that the Ct values were first affected before the changes in the quarantine policy affect the number of confirmed cases. Through these two time-series distribution graphs, it was assumed that it is possible to forecast the number of newly confirmed patients according to the distribution of Ct values.

Static estimation (simple linear regression model) and dynamic estimation (DLM) were used to perform forecasting of the actual occurrence of new confirmed cases using Ct values. First, as a result of the simple linear regression model, the forecasting was the best in the lag of 5-14 days. For forecasting through the simple linear regression model, the time difference with the highest predictive power differed by COVID-19 pandemic periods (11 days of the total period, 9 days of the first period, 5 days of the second period, 14 days of the third period, and 10 days of the fourth period). However, the results were likely to be false regression because it does not assume the normality of the time series. Also, the number of newly confirmed cases is difficult to forecast simply as a result of simultaneous static estimation, it was necessary to eliminate the effects of temporal fluctuations and make forecasts through dynamic estimation. After removing the effect of time fluctuations through DLM, both the third and fourth periods, which have no autocorrelation of the error term, were the best forecasted until the three-day time difference.

As a result of the forecasting through Rt in the same analysis method, it was always best forecasted at the 8-day time difference. And this study compared the forecasting of the number of newly confirmed cases through Ct values and Rt by specifying the time of Omicron spread. However, the analysis period was short for 30 days and it was hard to get significant results because the dominant species period of Omicron was from February 2022.

Conclusions This study confirmed the epidemiological utility of forecasting the newly confirmed cases by Ct values, which was rarely used in Korea. A COVID-19 forecast study was conducted using the domestic Ct values distribution at the population group level, including the time when the mutated virus appeared. It is the first domestic study to perform forecasts through Ct values using both static and dynamic estimations. This study revealed that Ct values can be used as a short-term (3 days) COVID-19 forecast indicator.

Based on the results of this study, if various follow-up studies are conducted by combining clinical and policy indicators, the epidemiological utility of Ct values will be further expanded to contribute to establishing health policies.

Keywords : Ct Values; Reproduction number; Dynamic estimation; Forecasting; COVID-19; Distributed Lag Model(DLM); Time lag;
Student Number : 2020-24281