

Original Article

# 식·의약 위해 감시체계(K-RISS)의 우선순위 평가를 위한 시계열 구조변화 기반 기준선 설정 모델 개발

진현정<sup>1\*</sup>, 허성윤<sup>1</sup>, 이현주<sup>2</sup>, 장보윤<sup>2</sup>

<sup>1</sup>중앙대학교 경영경제대학 경제학부, <sup>2</sup>캠아이넷(주)

## Development of a Baseline Setting Model Based on Time Series Structural Changes for Priority Assessment in the Korea Risk Information Surveillance System (K-RISS)

Hyun Joung Jin<sup>1\*</sup>, Seong-yoon Heo<sup>1</sup>, Hunjoo Lee<sup>2</sup>, and Boyoun Jang<sup>2</sup>

<sup>1</sup>Department of Economics, College of Business & Economics, Chung-Ang University, <sup>2</sup>CHEM. I. NET, Ltd.

### ABSTRACT

**Background:** The Korea Risk Information Surveillance System (K-RISS) was developed to enable the early detection of food and drug safety-related issues. Its goal is to deliver real-time risk indicators generated from ongoing food and drug risk monitoring. However, the existing K-RISS system suffers under several limitations.

**Objectives:** This study aims to augment K-RISS with more detailed indicators and establish a severity standard that takes into account structural changes in the daily time series of K-RISS values.

**Methods:** First, a Delphi survey was conducted to derive the required weights. Second, a control chart, commonly used in statistical process controls, was utilized to detect outliers and establish caution, attention, and serious levels for K-RISS values. Furthermore, Bai and Perron's method was employed to determine structural changes in K-RISS time series.

**Results:** The study incorporated 'closeness to life' and 'sustainability' indicators into K-RISS. It obtained the necessary weights through a survey of experts for integrating variables, combining indicators by data source, and aggregating sub K-RISS values. We defined caution, attention, and serious levels for both average and maximum values of daily K-RISS. Furthermore, when structural changes were detected, leading to significant variations in daily K-RISS values according to different periods, the study systematically verified these changes and derived respective severity levels for each period.

**Conclusions:** This study enhances the existing K-RISS system and introduces more advanced indicators. K-RISS is now more comprehensively equipped to serve as a risk warning index. The study has paved the way for an objective determination of whether the food safety risk index surpasses predefined thresholds through the application of severity levels.

**Key words:** K-RISS, food and drug safety, severity threshold, structural change, emerging risk

Received February 18, 2024

Revised April 1, 2024

Accepted April 17, 2024

### Highlights:

- Established a severity standard that takes into account structural changes in the time series of K-RISS.
- Survey for specialists was conducted to derive the required weights for K-RISS. Established cautious, attentive, and serious levels for K-RISS values.
- We systematically verified structural changes in K-RISS time series and derived respective severity levels for each period.

### \*Corresponding author:

Department of Economics, College of Business & Economics, Chung-Ang University, 84 Hekseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea  
 Tel: +82-2-820-5607  
 Fax: +82-2-3280-5582  
 E-mail: hyunjjin@cau.ac.kr

## 1. 서 론

정책적 모니터링의 수단으로 지수/지표의 활용이 다양한 분야에서 이루어지고 있다. 예를 들면, 한국은행(2022)의 뉴스 심리지수(News Sentiment Index, NSI),<sup>1)</sup> 대외경제정책연구원

(2018)의 금융불안지수(KIEP Financial Stress Index),<sup>2)</sup> 한국 소비자안전위원회(2018)의 소비자안전지표(Consumer Safety Indicators),<sup>3)</sup> 식품의약품안전처(2021)의 어린이 식생활 안전지수(Children's Dietary Life Safety Index)<sup>4)</sup> 등이 있다. 해외에는 유럽 지역의 공중 보건 및 환경 정책 지원을 목표로 국가별 수준과 지역 평



가로 구성된 WHO의 환경건강정보시스템(Environment and Health Information System, ENHIS),<sup>5)</sup> 국경 현장 사무소에서 환경 위해수준과 그에 따른 건강 영향을 모니터링하는 미국의 미국-멕시코 국경환경보건지표(Environmental Public Health Indicators),<sup>6)</sup> 인도주의적 위기 및 재난 위험에 처한 국가를 식별하기 위한 EU위원회의 INFORM 위험지수(INFORM Risk Index)<sup>7)</sup> 등이 있다.

국가적 차원에서 유해화학물질을 비롯한 신종위해요소(emerging risk)의 조기 발견은 국민의 건강을 위협하는 요소를 사전에 예방하고, 차단하거나, 경제적 피해를 최소화하기 위해 매우 중요하다. 신종위해요소를 발굴하기 위해서는 학술 문헌, 정부 보고서, 언론, 소셜네트워크, 소비자피해 신고 등 방대한 자료원으로부터 정보를 수집하여 잠재적 위해요소를 발굴·우선순위화하는 과정이 요구된다. 따라서, 이러한 광범위한 자료를 효과적으로 처리하기 위한 지식정보 시스템 구축이 필요하다.

식품의약품안전처는 2020년부터 식·의약 분야의 신종위해요소를 모니터링하고, 선제적으로 위해정보 이슈를 발굴하기 위해, 다양한 디지털·인공지능 기술을 접목한 식·의약 위해감시체계(Korea Risk Information Surveillance System, K-RISS)를 구축하고 있다. K-RISS는 2021년 식·의약품 등의 잠재이슈 업무지침을 통하여 일단위로 도출된 대량의 신종위해요소 후보에 대한 우선순위화를 위해 외부항목, 감정축발요인 지수를 정하고, 전자에는 심각성, 시급성, 파급력, 빈도수, 지속성을 후자에는 영유아 영향, 공포감, 생명과의 직결성을 각각 지수도출을 위한 평가 지표로 제시한 바 있다. 그러나, K-RISS의 기존 지표인 심각성(severity), 파급력(ripple effect), 지속성(persistence)만 가지고는 일상생활 속 제품의 중요도와 노출계수에 미치는 영향을 반영하기에 한계가 있다. 또한, 도출된 K-RISS 값에 대해 주의(cautious, CA), 경계(attentive, AT), 심각(serious, SE)을 의미하는 상대적 기준이나 지표 간 체계적인 가중치는 마련되어 있지 않은 상태였다.

이에 본 연구에서는 세부 지표를 추가하고, 일별로 도출되는 K-RISS 시계열 값의 구조변화를 감안하여 사안이 얼마나 중요한지 판단하는데 도움이 되는 상대적 기준치를 만드는 데 목적이 있다. 먼저, K-RISS의 기존지표 외에 생활밀접성(closeness to life)을 추가하고, 한 가지 자료원만 사용하던 지속성 지표를 모든 자료원에 대해서 도출하고, 변수나 지표를 합계하는데 필요한 가중치를 도출하고자 한다. 또한 도출된 K-RISS 시계열 값에 대해서 CA, AT, SE의 상대적 기준치를 설정하고, 일별 K-RISS 값이 특정 구간에서 많이 나타나거나 적게 나타나는 구조변화가 있는 경우 이를 검증하여 각 구간별로 다른 상대적 기준치를 도출하고자 한다.

## II. 재료 및 방법

### 1. K-RISS

K-RISS에서 신종위해인자를 발굴하기 위해 사용하는 자료원은 식품안전정보원(National Food Safety Information Service, NFSI)의 국내외위해정보시스템, 한국소비자원(Korea Consumer Agency)의 소비자위해감시시스템(CISS), 실험실정보관리시스템(Laboratory Information Management System, LIMS)의 수거검사 결과, 유럽식품사료신속경보(The Rapid Alert System for Food and Feeds, RASFF), 건강기능식품정보(Health Functional Food Information, HFFI)로 구성되었다. 각 자료원에 대해 평가요소로는 심각성(위해인자 변수, 발암성 등급과 생물학적 위해수준), 파급력(상위 100건의 다소비 품목), 빈도수에 따라 점수를 부여하였으며, 실험실정보관리시스템(LIMS)의 경우에는 지속성(직전 3년간 부적합 횟수)을 추가적으로 산출하였다.

### 2. 지표 및 변수의 조정

첫 번째 작업으로 기존 K-RISS 모델의 지표들과 측정 변수들을 다음과 같은 측면에서 수정하였다. 기존 모델에 일상생활 속에서 사용되는 제품의 중요도와 노출계수 측면에서 노출량에 미치는 영향을 반영하기 위해 생활밀접성을 나타내는 지표와 이에 대한 변수를 포함하였다. 이 과정에서 기존 모델에서 쓰인 변수들은 네 가지 지표에 하위변수로 재배치하였다. 예를 들어, 파급력을 측정하는데 사용되었던 '다소비 품목'은 본 연구에서 생활밀접성의 측정변수로 이동하였다. 한편, 지속성의 경우 LIMS 자료원에 한해서 도출하였는데, 다른 자료원에 대해서도 도출하기 위해 세부모형을 도입하였다. 수정된 K-RISS의 지표와 측정변수들이 Table 1에 제시되어 있다.

본 연구에서는 분석의 편의와 결과 제시의 명확성을 위해 여러 가지 자료원 중 가장 대표적인 위해관련 정보시스템인 식품안전정보원(NFSI)의 국내외위해정보시스템, 한국소비자원의 소비자위해감시시스템(CISS), 실험실정보관리시스템(LIMS)의 수거검사 결과를 이용하여 수정된 K-RISS 값을 도출하였다. NFSI는 국내외 뉴스기사와 유럽식품사료신속경보(RASFF)에서 제공하는 유럽 국가의 식품 및 사료에 대한 사건 데이터(자료 제외)가 제공된다. CISS에는 소비자위해감시시스템에서 제공하는 소비자신고 데이터가 제공되며, LIMS의 수거검사 결과는 수입/유통 식품에 대해 모니터링 검사를 통한 식품에 대한 부적합 판단 결과가 제공된다.

지표는 심각성, 파급력, 생활밀접성, 지속성 네 가지로 구성되는데, 이 중 심각성은 위해인자의 발암성 등급 변수와 위해인자의 생물학적 등급 변수로 구성하였다. 발암성 등급은 국제암연구소(International Agency for Research on Cancer, IARC), 미국산업위생사협회(American Conference of Governmental

**Table 1.** K-RISS indicators and measurement variables

Indicators	Variable	Data source		
		NFSI	CISS	LIMS
1. Severity	Carcinogenicity level of hazard	10	10	10
	Biological class of hazard	5	5	5
2. Ripple effect	Source of article on the issue	3	-	-
	Classification of information on the issue	-	3	-
	Non-conformity rate	-	-	1
3. Closeness to life	Grading based on purchase frequency of the product	5	5	5
	Grading based on quantity consumed of the product	5	5	5
4. Persistence	Degree of persistence of non-conformity	3	3	3

Industrial Hygienists, ACGIH), 미국 독성물질관리프로그램 (National Toxicology Program, NTP)에서 부여하는 점수를 이용하였다. IARC에서는 특정 요인이나 물질이 인간에게 암을 유발할 가능성을 평가하여 발암성 등급을 매기는데, 가장 위험성이 높은 그룹에서 가장 낮은 그룹 순으로 1군, 2A군, 2B군, 3군, 4군으로 분류하고 있다. ACGIH는 주로 직업 환경에서의 화학 물질 노출에 초점을 두고 가장 위험성이 높은 그룹에서 가장 낮은 그룹 순으로 A1, A2, A3, A4, A5로 분류한다. NTP는 주로 물질이 인간에게 발암성을 가지는지 여부에 초점을 두고 ‘알려진 인간 발암물질’ (Known Human Carcinogens, NHC)과 ‘인간에서 발암 가능성이 있는 물질’ (Reasonably Anticipated to be Human Carcinogens, RHC)로 구분한다. 본 연구에서는 IARC와 ACGIH의 경우 가장 위험성이 높은 등급은 5점 그리고 가장 낮은 등급은 1점을 각각 부여하였다. 그리고 NTP의 NHC는 의미상 IARC와 ACGIH의 위험성이 가장 높은 그룹에 해당되어 5점을 부여하고, RHC는 IARC와 ACGIH의 두 번째 위험성이 높은 그룹에 해당되어 4점을 부여하였다. 따라서 특정 요인이나 물질이 IARC, ACGIH, NTP 모두 혹은 세 가지 중 일부에서 평가를 받은 경우 2점에서 15점까지 점수가 분포된다. 본 연구에서는 총점 10점을 기준으로 표준화하였다.

생물학적 평가 기준은 식품공전, WHO (World Health Organization)의 항생제내성등급, 뉴질랜드 국가표준기구 (Standards New Zealand, Standard NZ)의 점수를 이용하였다. 식품공전에 제시된 식품에서 발견될 수 있는 유해 미생물을 대상으로 인간에게 미칠 수 있는 위험성의 정도를 그룹 1 (높은 위험성)에서 그룹 3 (낮은 위험성)까지 세 가지 단계로 분류하였다. 그룹 1은 모든 식품에 대해 강한 규격이 있는 미생물을 의미하고, 그룹 2는 환자, 어린이 등 취약 그룹이 주로 섭취하는 식품군에 대해 규격이 있는 미생물, 그리고 그룹 3은 모든 식품에서 충분히 검출될 수 있지만 그룹 1이나 그룹 2에 비해 상대적으로 덜 위험한 미생물을 의미한다. WHO

는 항생제를 내성 위험에 따라 ‘비판적으로 중요한 항생제’ (Critically Important Antimicrobials, CIA), ‘고도로 중요한 항생제’ (Highly Important Antimicrobials, HIA), ‘중요한 항생제’ (Important Antimicrobials, IA)로 분류하고 있다. Standard NZ는 미생물(바이러스, 박테리아, 곰팡이, 기생충 등)의 위험성과 관련하여 위험그룹 1 (Low Risk), 위험그룹 2 (Moderate Risk), 위험그룹 3 (High Risk), 위험그룹 4 (Very High Risk)로 분류하고 있다. 본 연구에서는 식품공전 분류에서 1등급에 최고점수 3점을 부여하고, 다음으로 2등급과 3등급에 각각 2점과 1점을 부여하였다. WHO의 항생제내성등급은 CIA, HIA, IA에 대해서 각각 3점, 2점, 1점을 부여하였다. 그리고 Standard NZ의 경우 위험그룹 1은 위험성이 없는 경우에 해당되기 때문에 0점을 부여하고 위험그룹 2부터 4까지 각각 1점에서 3점을 부여하였다. 따라서 특정 요인이나 물질이 세 가지 모두 혹은 세 가지 중 일부에서 평가를 받은 경우 3점에서 9점까지 점수가 분포된다. 본 연구에서는 총점 5점을 기준으로 표준화하였다. 예를 들어 심각성 지표에 있어서 어떤 물질이 IARC에서 3점을 받고, 다른 발암성 등급이 없는 경우, 그리고 식품공전에서 3등급과 Standard NZ에서 위험그룹 4로 분류된 경우, 발암성 등급 점수는 6점 그리고 생물학적 등급 점수는 최고점수인 5점이 된다.

파급력은 자료원의 특성에 맞추어 NFSI의 경우 해당 이슈의 기사 출처, CISS는 해당 이슈의 정보 구분, LIMS는 부적합 판정률로 구성하였다. 파급력의 총점은 LIMS의 수거검사 결과를 제외하고는 각각 최고 3점이 되도록 설정하였다. 예를 들어, 해당이슈의 기사출처와 정보 구분 같은 경우는 정부, 공공기관, 소비자단체는 3점, 학계, 산업계는 2점, 박람회, 온라인쇼핑몰은 1점, 기타는 0점으로 부여된다(언론계는 구독자 수를 고려하여 0~3점 부여). 한편, LIMS의 수거검사결과와 변수인 부적합 판정률은 부적합률/기준치-1로 산출한다. 예를 들어, 2020년 4월 23일 총 97,574건의 검사 중 32건이 부적합 판정을 받았고, 기준값(최근 3년간의 부적합률의 평균값)이 0.02%일 때,

2020년 4월 23일의 부적합률은 0.033%이라면 최종 기준 대비 부적합 판정률은 0.65가 된다. 지표 하위변수의 직관적인 이해도와 계산의 편의를 높이기 위해 NFSI와 CISS는 3점을 그리고 LIMS는 1점을 원점수로 하였다. 이 세 가지 변수들을 이용하여 평균값을 도출하는 데 있어서 모두 100으로 환산하여 단위를 통일하는 과정을 거쳤다.

생활밀접성은 일상생활 속에서 물질 검출 시 국민들의 우려와 노출 가능성 등 파급력이 높은 다소비 제품(예, 쌀 등) 및 다빈도 제품(예, 대파 등)의 등급변수로 구성하였다. 객관성 확보를 위하여 식품, 화장품, 위생용품 각각에 대하여 국민건강영양조사(제5기~제8기, 질병관리청), 화장품 위해평가를 위한 국내 화장품 사용량 조사(2017, 식품의약품안전처), 위생용품 생산실적 통계(2022, 식품의약품안전처) 등을 바탕으로 등급 기준을 마련하였다. 다소비 등급은 모든 자료원에 대해서 상위 100위에 포함되는 다소비 품목을 1~4분위로 구분하여 2점에서 최대 5점까지 부여하였다. 추가로 100위 아래의 품목은 1점을 부여하였다. 또한 다빈도 소비 등급을 포함하였는데, 이는 다소비 등급이 주는 정보인 얼마나 많이 소비하는가와는 별도로 소비자들은 자주 소비하는 품목에 대해 상대적으로 파급력이나 심각성을 크게 느낄 가능성이 있기 때문이다. 이 역시 상위 100위의 다빈도 품목을 1~4분위로 구분하여 2점에서 최대 5점까지 부여하고, 100위 외 품목은 1점으로 처리하였다.

한편, 마지막 지표인 지속성은 매년 정기적인 소비자 불만 접수 및 물질 모니터링을 통하여 과거 이력을 가지고 있는 NFSI, CISS 및 LIMS 자료원을 대상으로 제품과 물질별 통계분석을 통하여 평균치를 산출하고, 당해에 발생한 제품 및 물질의 발생빈도와 비교하여 평균치를 상회하는 정도로써 등급 기준을 마련하였다. NFSI와 CISS는 지난 3년간 연도별로 각각 해당 제품과 물질이 전년도 심각성 점수 평균을 넘어섰느냐를 기준으로 산출하였다(연도별 1점, 총 3점). 예를 들어, 2019년 기준으로 2018년과 2017년에 해당 제품 혹은 물질이 심각성 점수 평균을 넘어섰다는 것은 해당 피해자가 전년도에도 특정 제품 혹은 물질에 나타났었다는 것을 의미하는 동시에, 이 피해자가 중요하게 다루어져야 하는 것을 의미한다. 따라서 각 연도에 전년도 심각성 평균 점수를 넘어선 경우 1, 그렇지 않은 경우 0을 부여하였다. LIMS는 지난 3년 평균 부적합률 기반으로 산출하였다. 예를 들어, 2019년 기준으로 전년도인 2018년과 전년도인 2017에 대해서 부적합 추이를 관찰하여 점수를 부여하며, 이에 최대 3점을 가지게 된다.

### 3. 전체 지수 도출

K-RISS의 수정에 있어서 두 번째 작업은 전체 지수 도출 모형을 만드는 작업이다. 각 지표의 하위 변수를 합하여 하나의 지표를 만들거나 개별 지표들의 값을 합하여 하나의 K-RISS 지표를 도출하는 경우 각 셀의 원점수를 모두 100점 만점이 되도록

설정하였다. 변수에 가중치를 적용하기 위해서는 이들의 단위를 통일하는 것이 필요하다. 이를 위해 모든 변수들의 값을 100이 되도록 하였다. 이 과정에서 일부 변수는 소수점이 발생하게 되는데, 해당 변수의 모든 사안에 대해서 동일하게 적용되기 때문에 특정 요인이나 물질만 불리하게 혹은 유리하게 점수를 받거나 하지 않게 된다. 예컨대, 심각성에서 발암성 등급은 0~10점을 부여받기 때문에 10을 곱하여 최대 100이 되도록 하고, 생물학적 등급은 0~5점을 가지기 때문에 20을 곱하여 최대 100이 되도록 하였다. 이 둘을 합쳐서 각 자료원별 심각성 지표를 도출하는 경우 도출된 가중치를 적용하게 된다. 다음으로 세 가지 자료원을 이용하여 각 날짜의 심각성 지표를 도출하는 경우 다음과 같은 식에 바탕을 둔다.

$$Seris_t = w_{fsi} Seris_{fsi,t} + w_{cis} Seris_{cis,t} + w_{lms} Seris_{lms,t} \quad (1)$$

여기서  $Seris_t$ 는 특정 날짜의 심각성 지표 값을 의미한다. 그리고  $Seris_{k,t}$ 처럼 하첨자  $k$ 가 있는 경우 이는 각 자료원별 심각성 값을 의미한다. 여기서  $k$ 는  $fsi$ ,  $cis$ ,  $lms$ 로 표시되는데, 이는 각각의 자료원인 NFSI, CISS, LIMS를 의미한다.  $w_k$ 는 전체 자료원에 걸쳐 심각성 지표 도출 시 필요한 각 자료원별 가중치를 의미한다. 이렇게 각각의 지표별로 개별 자료원으로부터 도출된 값을 가중치를 이용하여 총점 100이 되도록 도출할 수 있다. 가중치를 이용하는 경우 자료원별, 지표별 변수값들을 같은 단위로 맞추어야 하기 때문에, 각 변수들의 총점을 100으로 다시 환산하였다. 심각성 지표 이외에 파급력(Ripple), 생활밀접성(Life), 그리고 지속성(Cont) 역시 같은 방식으로 구할 수 있으며, 비슷한 기호로 표시한다.

개별 지표에 대해서 자료원들의 가중치를 이용하여 합산하는 것이 Table 1에서 가로 합계를 의미하며, 각 자료원에 대해서 네 개의 지표를 합하여 자료원별 K-RISS 값을 도출하는 것이 세로 합계를 의미한다. 예를 들어, 특정 날짜에 대한 NFSI 자료원에 대한  $K-RISS_{fsi,t}$ 를 구하고자 한다면, 이 자료원에 대한 K-RISS 값 도출은 다음 식에 의거한다.

$$K-RISS_{fsi,t} = w_{seris} Seris_{fsi,t} + w_{ripple} Ripple_{fsi,t} + w_{life} Life_{fsi,t} + w_{cont} Cont_{fsi,t} \quad (2)$$

여기서  $w_{seris}$ ,  $w_{ripple}$ ,  $w_{life}$ ,  $w_{cont}$ 는 각각 심각성, 파급력, 생활밀접성, 지속성의 가중치를 의미한다. 이 가중치를 이용하여 각 자료원별 K-RISS 값이 총점 100이 되도록 도출할 수 있다. 여기서  $Ripple_t$ 는 특정 날짜의 모든 자료원에 대해 합산된 파급력을 의미하며,  $Ripple_{fsi,t}$ 는 특정 날짜의 NFSI 자료원에 대해서 구해진 파급력을 나타낸다. 다른 지표에 대해서도 그 표시와 해석이 이에 준한다.

마지막으로 전체 자료원과 전체 하위지표를 합하여 K-RISS를 만들 수 있는데, 이는 두 가지 방향에 의해 접근할 수 있다.

첫째는, 식(1)에 의거하여 각 하위지표들을 모든 자료원들에 대해 구한 다음, 이 하위지표 값들을 세로로 합하는 방식이다. 이는 다음과 같은 식에 의해 구해진다.

$$K-RISS_t = w_{seris}Seris_t + w_{ripple}Ripple_t + w_{life}Life_t + w_{cont}Cont_t \quad (3)$$

한편 식(2)에 구해진 각각의 자료원별 K-RISS 값들을 가로로 합하여도 구할 수 있는데, 다음과 같은 식에 의거한다.

$$K-RISS_t = w_{fsi}K-RISS_{fsi,t} + w_{cis}K-RISS_{cis,t} + w_{lms}K-RISS_{lms,t} \quad (4)$$

#### 4. 가중치 설정

세 번째 작업은 변수를 이용하여 지표를 도출하거나, Table 1의 각 지표들을 세로 혹은 가로로 합하여 각 자료원별 K-RISS나 전체 K-RISS를 도출하는 데 있어서 필요한 가중치를 도출하는 것이다. 심각성의 두 변수인 위해인자의 발암성과 생물학적 등급은 우리나라 질병 중 암 관련 질병의 비용과 식품관련 세균 및 미생물관련 질병의 비용의 비율을 이용하여 가중치를 도출하였다. 국민건강보험공단과 건강보험심사평가원의 「건강보험통계」 중 ‘질병소분류별 다빈도 상병 급여현황’ 2020년 자료를 이용하였다.<sup>8)</sup> 이 자료에는 진료실인원, 내원일수, 급여일수, 진료비, 급여비가 제시되어 있다. 2020년 이전 자료도 포함하는 것이 바람직하나, 매년 다빈도 상병의 순위에 큰 변동이 없기 때문에 분석시점에서 가장 최근인 2020년 자료를 이용하였다. 관련 질병이 사회적으로 어느 정도 비용을 초래하였는지가 중요하기 때문에 자료에 제시된 여러 변수들 중 진료비 항목을 이용하였다.

전체 자료원에 대해 파급력 값을 도출하는 경우 각 자료원의 변수에 대한 가중치를 어떻게 주어야 할지에 대해서 전문가 조사를 통하여 확정하였다. 같은 방식으로 생활밀접성의 다빈도 등급과 다소비 등급의 가중치, 각 자료원별 합산 시 필요한 가중치, 그리고 각 하위지표를 합하여 K-RISS를 만드는 경우에 필요한 가중치 또한 전문가 조사를 통하여 도출하였다. 전문가 조사는 100점 만점 고정총합법을 사용하였다. 즉, 가중치가 필요한 항목에 대해 세부항목들을 나누고, 세부항목 전체의 합계가 100점 만점이 되도록 제시한 후 전문가 본인이 생각하는 가중치 비중을 적도록 하였다. 파급력, 생활밀접성 지표 내 가중치 설정, 자료원 간 가중치 설정, 그리고 네 가지 지표 간 가중치 설정을 위해 네 개의 고정총합 질문을 제시하였다. 조사는 2022년 9월 21에서 10월 11일까지 이루어졌으며, 식품안전 및 영양 그리고 의학 전문가 20명에게 조사표 배포 후 수거하는 방식으로 자료를 확보하였다.

#### 5. 위험도 수준 설정

네 번째 단계로 도출된 각 날짜별 K-RISS 값 혹은 지표값의

평균값과 최댓값에 대해서 주의, 경계, 심각 수준을 설정하였다. 어느 정도 값이 나타나야 이러한 기준을 넘어서는 것으로 간주할 것인가에 대해 기준이 없기 때문에, 데이터의 이상점 및 변화점(change point) 탐지에 주로 사용되는 통계적 공정관리(statistical process control, SPC) 도구인 관리도(control chart)를 사용하여 상대적 기준치를 설정하였다.<sup>9,10)</sup> 설정된 수준에 의하여 만일 심각(SE)을 넘어가는 값이 나오지 않는 경우, 이는 정부가 주의를 기울여야 하는 사안을 하나도 걸러내지 못한다는 것을 의미한다. 이에 보정을 통하여 기준점을 넘어서는 값이 적절한 수준에서 나오도록 조정할 필요가 있다. 반대로 너무 많이 나오는 경우 정부에서 모든 사안에 대해서 주의를 기울여 살펴봐야 하는 것을 의미하기 때문에, 이 역시 적절한 보정이 필요하다. 따라서 정부에서 관심을 가지고 모니터링해야 하는 사안이 적절한 숫자가 되도록 조정하는 것이 바람직하다.

관리도는 품질 특성치의 값이 적절한 관리상태라면 관리한 계선 내에 존재하도록 설정하는 데서 시작된다. 따라서, 시간이 지남에 따라 특성치 값의 상태가 어떻게 변화하는지 시각적으로도 쉽게 파악할 수 있기 때문에, 많은 응용 분야에서 자주 사용되고 있다.<sup>10)</sup> 일반적으로 연속적으로 관측되는 일변량 혹은 다변량 데이터에 대하여 서로 독립적으로 일정한 분포를 따른다고 가정하고, 이들 분포를 특정짓는 모수의 변화 유무를 검정함으로써 이상점 또는 변화점을 탐지하게 된다.

관리도는 관측되는 데이터를 활용하는 방법에 따라 크게 두 가지로 나누어진다. 첫 번째는 현재 데이터만 이용하여 검정하는 경우로, 대표적인 방법론은 Shewhart 관리도가 있다.<sup>9)</sup> 두 번째는 현재뿐 아니라 과거 데이터를 동시에 고려하여 검정하는 경우로, 대표적인 방법론은 지수가중이동평균(exponentially weighted moving average, EWMA) 관리도<sup>11)</sup>가 있다. Shewhart 관리도는 모평균이 크게 변화할 때, EWMA 관리도는 모평균의 작은 변화에 대하여 탐지 성능이 뛰어나다고 알려져 있다.<sup>12)</sup> Cho 등(2012)<sup>13)</sup>은 기업의 대형장치 공정에서 30분 간격으로 1개월간 수집된 온도 데이터를 이용하여 EWMA 관리도와 Shewhart 관리도의 모니터링 효율성을 비교하였다. 공정 데이터의 큰 변화는 두 관리도 모두에서 검출되었지만, 근접한 점들 간 미세한 변화에 대해서는 EWMA 관리도에서만 검출됨을 확인하였다. 본 연구에서는 현재 발생하는 K-RISS 자료의 상한선을 설정하는 작업이기 때문에 Shewhart 관리도를 이용하고자 한다. Shewhart 관리도는 구현하기가 간단하며, 이상 원인에 대한 평균의 변동이 큰 경우 발생 유무를 빨리 감지할 수 있다는 장점이 있다.<sup>14)</sup>

Shewhart 관리도의 기본 모형은 다음과 같다. 개별 관측치  $x_1, x_2, \dots, x_k$ 가 연속적으로 관측되며 관리상태일 때 이들 관측값은 모평균이  $\mu_0$ , 모분산이  $\sigma_0^2$ 인 정규분포를 따른다고 가정한다. 이때  $\mu_0$ 와  $\sigma_0^2$ 이 알려진 경우를 가정하고, 새롭게 관측되는 데이터에 대하여 모평균의 변화를 검정한다. 이 검정은 다음

과 같은 관리통계량( $X$ )으로 이루어진다.

$$\text{관리상한선(upper control limit, UCL)} = \mu_X + L_X \sigma_X \quad (5)$$

$$\text{중심선(center line, CL)} = \mu_X \quad (6)$$

$$\text{관리하한선(lower control limit, LCL)} = \mu_X - L_X \sigma_X \quad (7)$$

여기서 관리상한선(UCL)은 공정이 제어가능한 최대 한계를 나타내며, 중심선(CL)은 공정의 평균값, 그리고 관리하한선(LCL)은 공정이 제어가능한 최소 한계를 나타낸다.  $L_X$ 는 관리한계선의 폭을 결정하며, 관리도를 설계할 때 결정해야 하는 관리모수이다. 관리도의 가로축은 시간을, 세로축은 데이터 단위를 나타내며, 샘플의 평균을 나타낸 중앙선(central line)과 관리 한계(control limits-UCL, LCL)를 가로축과 평행으로 나타낸다. 통계량  $X$ 가 관리한계선 내에 존재하면, 즉  $X \in (LCL, UCL)$ 이면, 관리상태( $\mu = \mu_0$ )라고 판정한다. 반대로, 통계량  $X$ 가 관리한계선 밖에 존재하면, 즉  $X \notin (LCL, UCL)$ 이면, 이상상태( $\mu > \mu_0$ )라고 판정한다. 식(5)~(7)을 통해 공정의 변동을 정량적으로 측정하고 제어하는 데 필요한 기준을 설정할 수 있다. 이 수식들은 공정이 적절한 상태에 있는지, 아니면 조정이 필요한지를 판별하는 데 있어서 핵심적인 역할을 한다. UCL과 LCL은 공정에서 관측된 데이터가 허용하는 최대 및 최소 범위를 의미하며, 일반적으로 3시그마를 사용한다. 이는 정규분포가 정 하에 적절한 상태의 공정에서 관측된 데이터의 약 99.73%가 이 범위 내에 놓인다는 것을 의미한다. 따라서, UCL 이상이나 LCL 이하의 값은 공정에 문제가 있음을 나타내는 신호로 간주된다. 중심선(CL)은 공정의 평균적인 성능을 의미하는데, 공정이 안정적이고 예측가능할 때의 기대치를 나타낸다. 요약하자면, 위 식들은 변동성 모니터링에 기반을 두고 공정이 제어 상태에 있는지를 효과적으로 판단하고, 필요한 경우 적절한 조치를 취할 수 있도록 기준치를 제공한다. 정규분포와 같은 통계적 가정이 현실적인 공정 데이터에 적용 가능하다는 전제 하에, 이들은 공정의 자연적인 변동을 인정하면서도, 비정상적인 변동을 식별할 수 있는 수단을 제공하는 것이다.

관리도의 성능을 평가할 때는 런길이(run length, RL)를 사용한다. 런길이는 관리한계선을 처음으로 벗어나 이상 신호가 발생할 때까지 관측한 표본의 개수를 의미한다. 관리상태인 경우 관리통계량  $X_t$ 가 관리한계선 밖에서 관측될 확률은 아래 식을 만족한다

$$P(X_t \notin (LCL_X, UCL_X) \mid \mu = \mu_0) = \alpha, t=1, 2, \dots, \quad (8)$$

관리상태 하에서  $RL_X$ 의 분포는 모수가  $\alpha$ 인 기하분포를 따르고, 이때의 기댓값을 평균 런길이(average run length, ARL)라

고 하며,  $ARL_0$ 로 표시한다.

$$ARL_0 = \sum_{k=1}^{\infty} k(1-\alpha)^{k-1} \alpha = \frac{1}{\alpha} \quad (9)$$

$X$  관리도의  $ARL_0$ 는 제1종 오류와 관계가 있고, 관리도의 성능을 나타낸다. 일반적으로 관리한계선은 통계적 유의성 검정 절차와 마찬가지로 미리 정한 수준의  $ARL_0$ 를 만족할 수 있도록 설정된다. 예를 들어,  $L_X$ 는 보통 3으로 주어지는데, 이것은  $\alpha=0.27\%$ 에 해당하며,  $ARL_0=370$ 임을 의미한다. 즉,  $ARL_0=370$ 을 만족하는  $X$  관리도는 관리모수가  $L_X=3$ 으로 설정된다고 할 수 있다. 식(5)~(7)에 의해 어느 공정이 관리범위 내에 있고 공정 통계가 정상이라면, 보통 99.73%의 데이터 값들이 UCL과 LCL 사이에 있게 된다. 따라서 이 개념을 적용하는 경우,  $(1-0.9973)/2$ 의 값이 UCL에 해당된다. 이 값을 SE 기준으로 설정하고, 이 값의 아래인 2시그마를 AT, 1시그마를 CA로 설정할 수 있다.

## 6. 시계열 구조변화 검출

날짜별로 K-RISS 값이 지속적으로 도출됨에 따라 앞에서 구한 CA, AT, SE 수준을 넘어가는 값들이 시기별로 많이 혹은 적게 나타날 수 있다. 이 경우 시계열 지표값의 기준을 수정해야 하는지에 대한 판단이 필요하다. 이에 본 연구에서는 시계열 구조변화모형을 활용하여 구조변화 시점을 중심으로 새로운 평균값과 위험도 수준을 도출하였다.

시계열자료의 구조변화를 추정하는 방법론은 통계학과 계량경제학 분야에서 연구가 많이 진행되어 왔다. 이 연구들 대부분은 특정 시점을 중심으로 전·후에 통계적으로 유의한 변화가 있었는지를 찾아내는데 중점을 두고 있다.<sup>15,16)</sup> 이러한 방법론들은 최근에 더욱 발전하여 미지의 구조변화의 시점까지 찾아내는 모형이 개발되었다. Bai와 Perron (1998, 2003)<sup>17,18)</sup>은 Liu 등(1997)<sup>19)</sup>의 연구를 발전시켜 시계열자료에 있어서 자기상관이나 이분산이 존재하는 경우에도 미지의 다중 구조변화를 찾아내는 모형을 제시하였다. 본 연구에서는 Bai와 Perron 방법론을 이용하여 특정 시점을 중심으로 전·후기에 통계적으로 유의한 변화가 있었는지 검증하고자 한다.

먼저 아래 제시된 바와 같이  $m$ 개의 구조변화, 따라서  $m+1$ 개의 하부시기를 가진 다중선형회귀식을 가정한다.

$$y_t = \beta x'_t + \theta z'_t + e_t, \quad t = \tau_{j-1} + 1, \dots, \tau_j, j=1, \dots, m+1 \quad (10)$$

여기서  $y_t$ 는  $t$ 기에 측정된 종속변수를 의미하며,  $x_t$  ( $p \times 1$ )와  $z_t$  ( $q \times 1$ )는 독립변수를 나타낸다.  $\beta$ 와  $\theta_j$  ( $j=1, \dots, m+1$ )는  $x_t$ 와  $z_t$ 에 해당되는 추정계수를 의미한다.  $x_t$ 는 전 기간에 걸쳐서 구조변화를 보이지 않는 독립변수를 의미하며,  $z_t$ 는 하부시기별로

구조변화를 보이는 독립변수를 의미한다.  $e_t$ 는 오차항을 나타낸다.

이 모형에서는 구조변화시점(break point)을 나타내는  $\tau_1, \dots, \tau_m$ 은 알려지지 않았다고 가정하고, 구조변화시점과 더불어 세부 시기별로 계수를 추정하게 된다. Bai-Perron 모형은 '부분구조변화모형'으로 불리우는데, 이는 독립변수 중 일부는 구조적으로 변하지 않아 전 기간에 걸쳐서 계수가 하나만 추정되고, 일부는 구조변화를 기점으로 하부시기 만큼의 숫자에 해당하는 계수가 추정되기 때문이다. Bai-Perron 모형은 OLS (ordinary least squares)의 원리에 바탕을 두고 있다. 또한 이 모형은 OLS 뿐만 아니라 자기회귀(auto-regressive)와 추세모형에도 적합시켜 구조변화를 찾아낼 수 있다는 장점이 있다. 또한 특정 시점을 기준으로 나누어진 시계열의 전·후기 오차항의 분포가 다르더라도 구조변화를 비교적 정확하게 찾아낼 수 있다는 특징이 있다.

$m$ 개의 구조변화시점( $\tau_1, \dots, \tau_m$ )을  $\{\tau_j\}$ 로 나타내고, 이에 따라  $m+1$ 개의 하부시기별로 추정계수인  $\beta$ 와  $\theta_j(j=1, \dots, m+1)$ 가 존재한다면, 다음과 같은 최소자승법에 의하여 계수의 추정치를 구할 수 있다.

$$\sum_{j=1}^{m+1} \sum_{t=\tau_{j-1}+1}^{\tau_j} [y_t - x'_t\beta - z'_t\theta_j]^2 \quad (11)$$

여기서  $\hat{\beta}(\{T_j\})$ 과  $\hat{\theta}(\{T_j\})$ 을  $m+1$ 시기의 추정치라고 하고, 이 추정치들을 목적함수에 대입하면, 오차자승의 합인 ST ( $\tau_1, \dots, \tau_{m+1}$ )를 이용하여 구조변화시점( $\hat{\tau}_1, \dots, \hat{\tau}_{m+1}$ )의 추정치들을 구할 수 있다. 따라서 구조변화 추정치는 전체목적함수를 최소화하는 값이며, 전체식의 계수 추정치는  $m+1$ 개 하부시기의 추정치와 동일하게 된다.

### III. 결 과

#### 1. 가중치 도출

지표 내 변수 간 가중치, 자료원 간 가중치, 그리고 지표 간 가중치 설정을 위한 전문가 조사 결과가 Table 2에 제시되어 있다. 결과에 대한 해석은 다음과 같다. 파급력의 경우, NFSI의 기사 출처, CISS의 정보 구분, LIMS의 부적합 판정률에 대해서 각각 0.403, 0.122, 0.476으로 나타났다. 이를 다 합치면 1이 된다. 따라서 특정 사안에 대한 파급력을 도출할 때, 각 변수값을 도출한 다음 도출된 가중치를 적용하여 합하면 된다. 만일 특정 변수 값이 0일 경우에는 해당 변수의 가중치를 빼고 다른 두 가지의 가중치를 1로 만든 다음 두 개 가중치의 크기를 감안하여 새로운 가중치를 만들게 된다. 생활밀접성의 경우에는 다빈도가 0.439 그리고 다소비가 0.561로 나타났다. K-RISS 나 특정 지표에 대해서 각 자료원 값들을 합산하는 경우 사용될 가중치는 NFSI, CISS, LIMS 각각 0.417, 0.118, 0.464로 나타났다. 최종적으로 심각성, 파급력, 생활밀접성, 지속성 이 네 가지에 대한 가중치는 각각 0.372, 0.242, 0.228, 0.158로 나타났다. 한편 심각성의 경우, 암 관련 질병 비용은 입원 및 외래환자 2020년 기준 4,895,978백만 원으로 나타났으며, 식품과 관련된 생물학적 위해인자에 의한 비용은 입원 및 외래환자 2020년 기준 3,961,707백만 원으로 나타났다. 따라서 발암성 등급과 생물학적 평가 등급의 가중치는 각각 0.553과 0.447로 도출되었다.

#### 2. 지표 값 도출

결과의 일부가 예로써 Table 3에 제시되어 있다. 표에 제시된 값을 보면, 각 자료원별로 네 가지 지표의 평균값과 최댓값이 제시되어 있다. 예를 들어, NFSI의 평균값은 NFSI에서 제시된 사안에 대해서 발암성 등급과 생물학적 등급 변수 평균값을 각각 100으로 환산한 후 0.533과 0.447의 가중치를 이용하여 총점 100이 되도록 도출하였다. 파급력의 경우에는 NFSI만

Table 2. Results of weight for variables and indicators

Indicators		Variable	NFSI	CISS	LIMS
Indicators	Weight		0.417	0.118	0.464
1. Severity	0.372	Carcinogenicity level of hazard	0.553	0.553	0.553
		Biological class of hazard	0.447	0.447	0.447
2. Ripple effect	0.242	Source of article on the issue	0.403	-	-
		Classification of information on the issue	-	0.122	-
		Non-conformity rate	-	-	0.476
3. Closeness to life	0.228	Grading based on purchase frequency of the product	0.439	0.439	0.439
		Grading based on quantity consumed of the product	0.561	0.561	0.561
4. Persistence	0.158	Degree of persistence of non-conformity	1	1	1

Table 3. Example of results for the mean and max of daily K-RISS values

Date	Mean				Max			
	NFSI	CISS	LIMS	K-RISS	NFSI	CISS	LIMS	K-RISS
20200101	23.60	-	-	9.85	68.07	-	-	28.40
20200102	24.57	3.47	27.65	23.50	78.79	17.34	59.20	62.42
20200103	24.12	10.19	17.48	19.39	70.27	30.57	42.80	52.81
20200104	28.66	-	-	11.96	60.28	-	-	25.15
20200105	22.72	-	-	9.48	52.34	-	-	21.84
20200106	34.24	3.19	24.26	25.93	79.24	28.75	56.69	62.79
20200107	24.62	21.09	19.06	21.62	78.79	46.26	56.69	64.68
20200108	32.71	15.88	24.58	26.94	81.29	35.68	52.58	62.56
20200109	30.71	23.02	24.23	26.79	78.79	46.26	46.88	60.12
20200110	28.85	21.90	19.99	23.91	70.57	46.26	47.80	57.12
20200111	37.58	-	-	15.68	40.00	-	-	16.69
20200112	23.05	-	-	9.62	49.99	-	-	20.86
20200113	29.52	20.94	23.75	25.83	83.35	46.26	52.57	64.67
20200114	33.00	18.79	23.25	26.79	78.79	43.28	52.57	62.41
20200115	33.89	17.97	22.14	26.55	70.27	55.84	52.58	60.35
20200116	29.51	-	25.09	23.96	69.44	-	52.57	53.39
20200117	30.92	21.92	18.49	24.08	75.12	43.28	44.35	57.06
20200118	42.56	-	-	17.76	76.30	-	-	31.84
20200119	15.83	-	-	6.61	15.83	-	-	6.61
20200120	34.58	21.45	22.11	27.23	78.79	43.28	44.35	58.60
20200121	34.62	26.88	20.57	27.18	81.29	55.84	55.07	66.11
20200122	30.09	15.31	23.68	25.36	72.62	26.77	47.57	55.56
20200123	33.59	27.10	14.66	24.03	71.01	46.26	50.52	58.57

도출하는 경우 '정보 소스' 변수만 있기 때문에, 이 값을 그대로 100으로 환산하였다. 생활밀접성의 경우 역시 다빈도 등급과 다소비 등급의 평균값을 각각 100으로 환산한 후 0.439와 0.561의 가중치를 이용하여 총점 100이 되도록 도출하였다. 지속성의 경우 변수가 하나밖에 없기 때문에 부적합 지속 정도 평균값을 그대로 100으로 환산하였다. 다음 단계로 NFSI의 각 지표별 가중치를 이용하여 평균값에 대한  $K-RISS_{i,t}$ 를 도출하였다. 이 값은 각 지표를 100점 만점 기준으로 도출하였기 때문에, 가중치를 적용하는 경우 다른 변환 없이도 100점 만점을 갖는 값이 된다. 같은 방식으로 CISS와 LIMS에 대해서 평균값에 대한 K-RISS를 도출하였다. 이후 자료원별 가중치를 이용하여 전체 K-RISS 값을 도출하였다.

한편 일별 평균값의 오른쪽에는 최댓값에 대한 자료원별 하위 K-RISS와 전체 K-RISS 값이 제시되어 있다. 하루에 여러 사안이 발생할 수도 있기 때문에 일별 평균과 최댓값을 구하였다. 특정 날짜에 있어서 평균값이나 최댓값이 없을 수도 있는데, 이는 해당날짜에 세 가지 정보원 모두에 보고된 사안이 없음을 의미하며, 이 경우 일별 평균과 최댓값은 0이 된다. 세 가지 자

료원에 있어서 동일한 한 건의 보고만 있는 경우 평균값과 최댓값이 같아지게 된다. 한편, 특정 정보원에서만 사안이 보고되지 않은 경우 K-RISS를 도출하는 데 있어서 해당 자료원의 가중치는 0으로 설정하였다. CA, AT, SE 기준치를 도출하는 과정에서 일별 평균값 대신 최댓값을 사용하였는데, 이는 우리가 관심을 갖는 것은 특정 날짜에 발생한 사안들 중 K-RISS 값이 가장 크게 나온 값, 즉 정부에서 관심을 기울여야 하는 사안을 찾아내는 것이 목표이기 때문이다.

### 3. K-RISS 상한선 설정

Shewhart X 관리도 관리상한과 하한인 UCL과 LCL을 바탕으로 중앙선으로부터 상하로 3시그마를 SE로 설정하고, 이 값의 아래인 2시그마를 AT, 1시그마를 CA로 설정하였다. 이를 시험하기 위해 2020년 1월 1일부터 12월 31일의 세 자료원에 보고된 사안들을 이용하여 최댓값 K-RISS를 도출하였다. 세 가지 모든 자료원에 대해서 사안이 없는 날을 제외하게 되면 관측치는 303개가 된다. 이 경우 전체 평균은 36.60 그리고 표준편차는 16.20으로 나타났다. 이에 따라 CA는 52.69, AT는

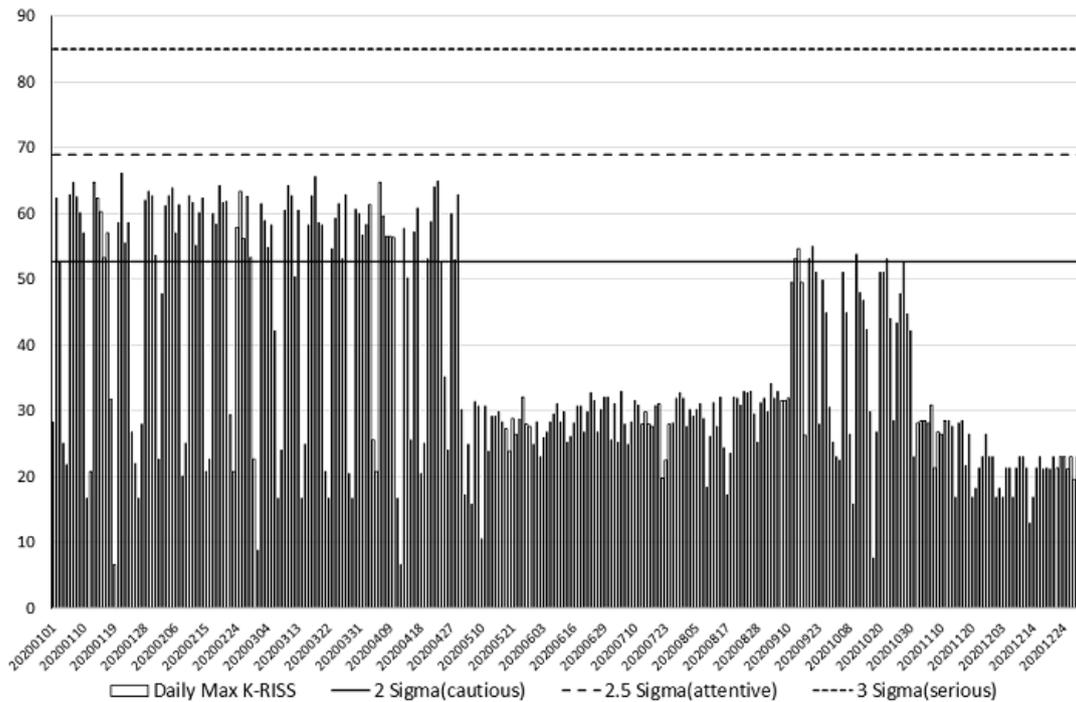


Fig. 1. Changes in daily maximum K-RISS values and UCL setting based on standard 3sigma in 2020

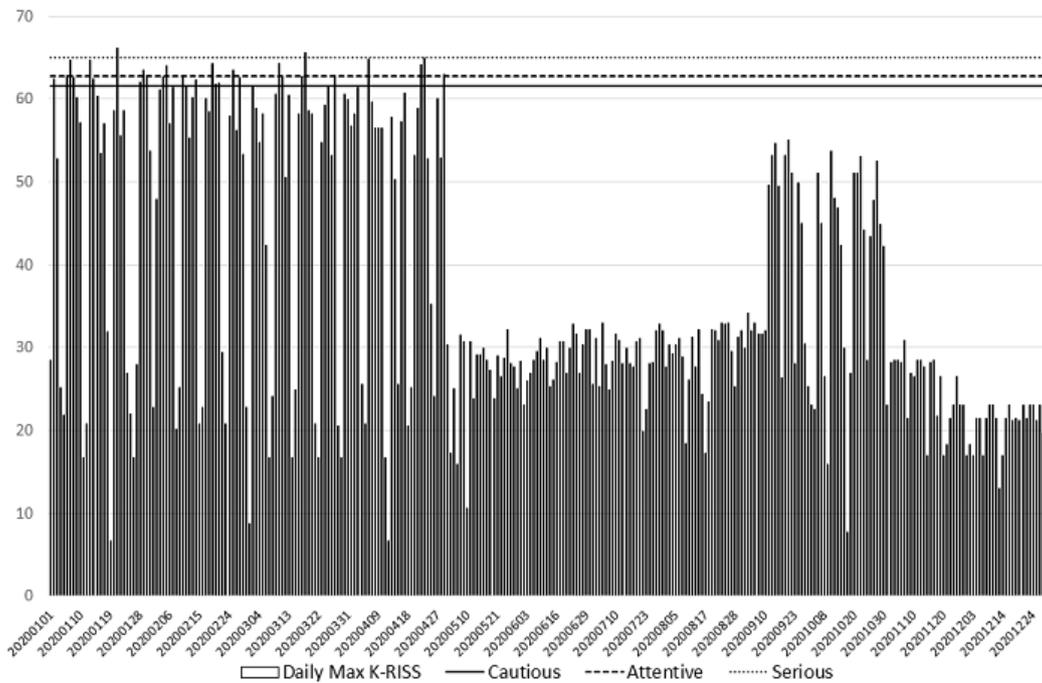


Fig. 2. Changes in daily maximum K-RISS values and UCL setting based on significance level in 2020

68.89, SE는 85.09로 도출된다. Fig. 1에 도출된 K-RISS의 일일 최댓값의 시계열자료의 모습이 제시되어 있다.

막대그래프로 제시된 값은 일 최댓값의 시계열 변화를 의미

한다. 그리고 가로로 제시된 선은 위에서부터 각각 SE, AT, CA를 나타낸다. 그림을 살펴보면, 시그마를 기준으로 이 기준치들을 설정하는 경우 CA를 넘어가는 값은 나타나지만, AT 이상을

넘어가는 값이 나타나지 않은 것을 알 수 있다. Shewhart가 보통 99.73%의 데이터 값들이 UCL과 LCL 사이에 있게 된다고 제시한 것을 감안하면,  $(1-0.9973)/2$ 의 값을 UCL로 적용하는 경우, 만일 관측치가 적어지게 되면 SE에 들어가는 관측치가 0 이나 1 이하가 될 수 있다. 이는 정책 당국이 관심을 가지고 지켜봐야 하는 사안이 하나도 도출되지 않을 가능성을 의미한다.

이 경우 SE에 포함되는 개수를 조절하기 위해 0.9973을 기준으로 하기보다는 기존의 신뢰구간인 0.99, 0.95, 0.90을 사용할 수 있다. 만일 이 구간에도 들어오는 관측치가 적은 경우에는 이 보다 구간 값을 더 낮출 수 있다. 본 연구에서는 0.99를 3시그마로, 0.95를 2시그마로, 0.90을 1시그마로 하여 기준치를 재설정하였다. 결과가 Fig. 2에 제시되어 있는데, 전보다 이 값을 넘어가는 값이 적절하게 나타나는 것을 알 수 있다.

#### 4. 구조변화 적용 후 재설정

Fig. 2를 보면, 시계열 값들이 시간이 지남에 따라 다른 모습을 가지는 것을 알 수 있다. 4월 29일까지는 다소 높은 값들이 지속되다가, 그 이후부터 평균값들이 전체적으로 하락한 것으로 나타난다. 따라서 이 시계열자료는 최소한 한 번 이상의 구조적 변화가 있음을 알 수 있다. 이는 전체 기간에 대해서 설정된 기준치에 의하면 5월 이후부터는 CA 이상의 값이 나타날 수 없음을 제시한다. 즉, 전체에 걸쳐서 기준치들을 한 번만 설정하게 되면, 어느 특정 기간에서는 이러한 선을 넘어가는 값이 전혀 나오지 않게 된다. 따라서 구조적 변화가 일어나서 평균값들이 전후로 큰 차이를 보인다면, 그 변화 시점을 기준으로 영역을 구분하는 것이 필요하다.

이에 시계열 자료의 구조적 변화가 통계적으로 유의한지를 Bai와 Perron 방법을 이용하여 테스트하였고, 결과가 Table 4에 제시되어 있다. 분석을 위해 구조적 변화의 최대치를 넘겨야 5로 하고, 유의수준을 일반적인 5%로 설정한 결과, Table 4에 제시된 바와 같은 결과를 도출하였다. 추정된 구조적 변화를 감안한 OLS (Ordinary Least Squares)를 이용하였으며, 독립변수는 따로 추가하지 않고 평균의 변화만 살펴보았다. 결과를 보면, 시계열 관측치를 기준으로 122번째 즉 4월 말을 기준으로 통계적으로 유의한 구조적 변화가 일어난 것을 알 수 있다.

**Table 4.** Results of Bai-Perron test for daily maximum K-RISS values

Period	Variable	Coefficient	t-statistic
1~121 (121 obs)	Constant	47.226 (1.241)	38.07
122~303 (182 obs)	Constant	29.372 (1.011)	29.04
R-squared	0.29	Adjusted R-squared	0.29
Akaike info criterion	8.071	Schwarz criterion	8.096

The total number of observations is 303. A structural change occurred in the 122nd, matching with April 29. Both AIC (Akaike info criterion) and SC (Schwarz criterion) are indicators that measure the goodness of fit of the model, and the smaller values of them indicate better model fitness.

위 결과는 최댓값 K-RISS 시계열자료에 있어서 한 번의 구조변화가 발생하였고, 이에 따라 두 개의 다른 영역으로 구분된다는 것을 의미한다. 첫 번째 구간은 1월 1일부터 4월 29일까지이며, 두 번째 구간은 4월 30일부터 12월 31일이 된다.

Table 5에 두 개의 구간으로 구분한 후 0.99, 0.95, 0.90을 이용하여 재조정된 CA, AT, SE가 제시되어 있다. 또한 Fig. 3과 Fig. 4에 구조변화를 기준으로 이전 시기와 이후 시기에 대한 최댓값 K-RISS 시계열과 CA, AT, SE가 제시되어 있다. 구조변화를 감안하지 않고 설정하게 되면 두 번째 구간에서 기준치들에 K-RISS 값이 하나도 포함되지 않지만, 구조변화를 감안하여 재설정하게 되면 각 구간에서도 이 기준치를 넘어가는 값들이 적절하게 나타나는 것을 알 수 있다.

### IV. 고찰 및 결론

K-RISS는 식·의약품의 주요 이슈에 대한 조기 감지를 목적으로 만들어진 인공지능 통합 위해정보감시관리시스템이다. 그러나 기존 K-RISS 시스템에는 생활밀접성을 나타는 지표가 포함되지 않았으며, 특정 날짜에 발생한 사안들의 K-RISS 값이 얼마나 중요하게 다루어져야 하는지에 대한 상대적 기준이 마련되지 않았다. 또한 도출되는 K-RISS 값이 특정 기간에는 전체적으로 높게 혹은 낮게 나타나는 구조적 변화를 감안하지 않았다는 한계가 있다.

이에 본 연구에서는 다음과 같은 측면에서 기존 K-RISS를 보완하였다. 첫째, 기존 K-RISS에 생활밀접성 지표를 추가하고,

**Table 5.** Settings of the point of cautious, attentive, and serious for each period based on structural change test results

	Whole period	Period 1	Period 2
Mean	36.502	47.095	29.556
Standard deviation	16.196	18.497	9.437
Cautious (0.90)	61.533	63.398	44.953
Attentive (0.95)	62.786	64.651	51.016
Serious (0.99)	64.973	65.992	54.756

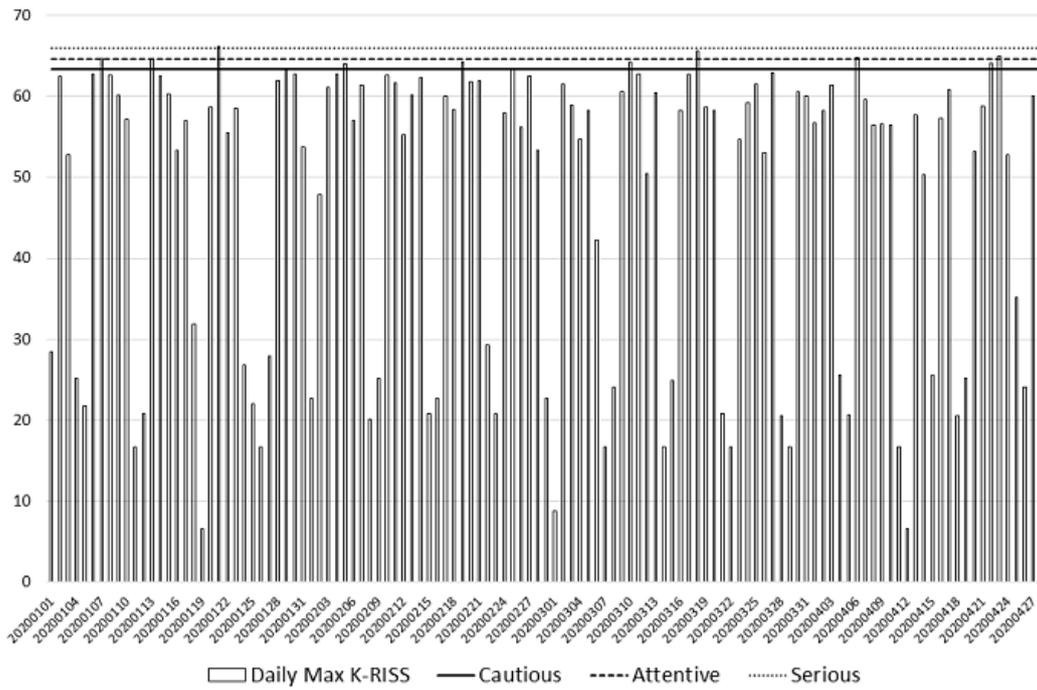


Fig. 3. Changes in daily maximum K-RISS values and UCL setting based on significance level (period 1 of structural change)

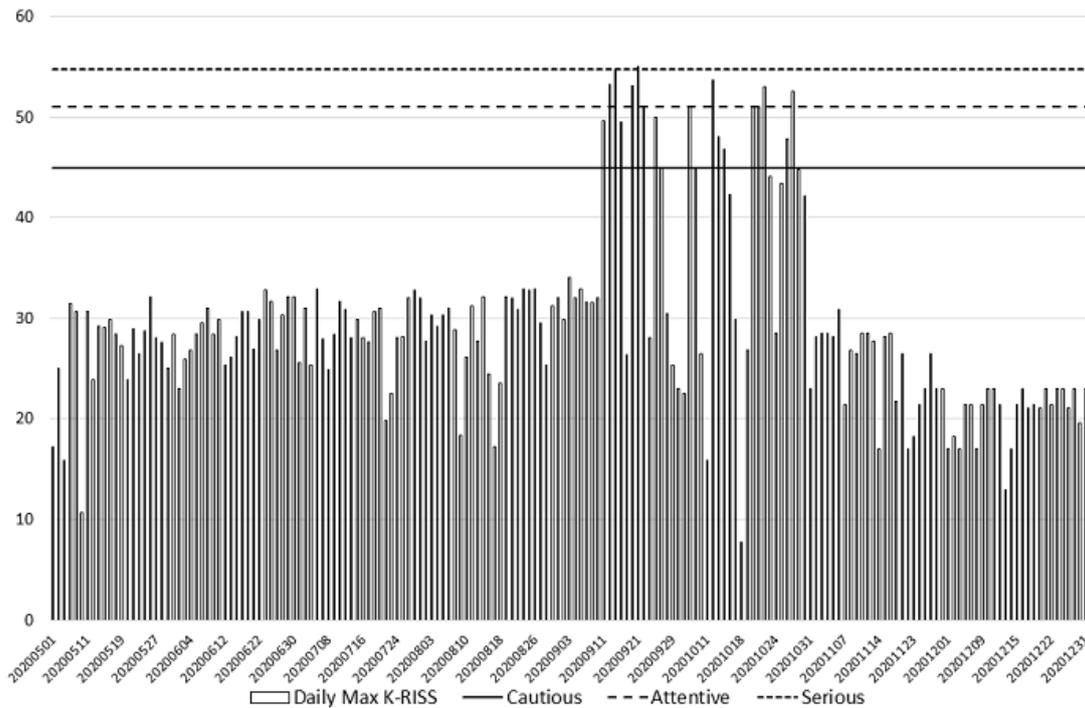


Fig. 4. Changes in daily maximum K-RISS values and UCL setting based on significance level (period 2 of structural change)

변수들을 지표들의 성격에 맞게 재배치하고, 지속성 지표를 모든 자료원에 대해서 도출하였다. 둘째, 변수를 통합하거나 지표를 통합하는 과정에서 필요한 가중치를 전문가 조사를 통하

여 도출하였다. 셋째, 도출된 일별 K-RISS 값의 평균값과 최댓값에 대해서 주의, 경계, 심각 기준치를 설정하였다. 이는 수집된 뉴스 및 자료를 바탕으로 도출된 식품안전위해지수가 특정

임계치를 넘어가는 경우 정부에서 주의를 기울여야 하는 사안으로 분류된다. 이를 위해 데이터의 이상점을 탐지하기 위해 통계적 공정관리에 사용되는 관리도를 이용하였다. 넷째, 일별 K-RISS 값이 특정 구간에서 너무 많이 혹은 적게 나타나는 구조변화가 있는 경우 이를 체계적으로 검증하여 각 구간별로 주의, 경계, 심각의 기준치를 각각 도출하였다. 이를 위해 시계열 구조변화모형을 활용하였다.

본 연구는 기존 K-RISS를 보완한다는 점에서 의의를 가진다. K-RISS에 생활밀접성 지표가 추가되고 모든 자료원에 대해서 지속성 지표가 계산됨으로써 기존의 심각성과 파급력 지표에 더하여 위해경보지수로서 필요한 항목들을 보다 체계적으로 갖추게 될 것이다. 또한 모든 사안에 대해서 주의를 기울이기보다는 설정된 주의, 경계, 심각 기준치를 넘어가는 사안에만 살펴볼 수 있는 체계를 갖출 수 있을 것으로 기대된다. Lee (2022)<sup>20</sup>는 2018년 3월 5일부터 2020년 3월 5일까지 국민청원 보건복지 범주에 올라온 청원에 대하여 주별 단위로 108개의 문서를 생성한 후 LDA 분석을 통한 토픽지수를 생성하고, 이 토픽지수를 개별 관측치로 하여 이에 대한 변화를 모니터링한 바 있다. 또한 일별로 도출되는 K-RISS 값에 구조변화가 있는 경우 이를 검증한 후 각 구간별로 상대적 기준치를 각각 도출하여 정부에서 살펴봐야 하는 사안의 숫자가 모든 기간에 적절하게 분포하도록 하는데 도움이 될 것이다.

본 연구는 위해조기경보체계인 K-RISS를 보완하였다는데 실용적 의의가 있다. 조정된 K-RISS 모형을 통해 유관기관 및 담당자들은 보다 효과적으로 실시간 식품위해를 모니터링할 수 있을 것으로 예상된다. 즉, 식품의약품 위해요소가 사회적으로 문제가 되기 전 미리 짐작할 수 있는 실마리를 도출하는데 도움이 될 것으로 기대된다.

한편, 본 연구는 다음과 같은 점에서 한계를 노출하고 있는 바 향후 연구에서 추가적인 분석이 필요하다고 판단된다. 첫째, 가중치 결정에 있어 전문가 조사를 이용하였으나, 응답자 표본의 숫자가 크지 않기 때문에 개별 전문가의 주관적 요소를 배제하는데 한계가 있다고 판단된다. 따라서 향후에는 표본의 숫자와 다양한 분야에 걸친 전문가 조사를 통해 가중치를 점검할 필요가 있다고 판단된다. 둘째, 본 연구의 목적은 식·의약 위해 감시체계(K-RISS)의 효과성을 높이기 위해, 세부 지표를 추가하고, 사안이 중요성을 선별하는 상대적 기준치를 설정하는 것이다. 현재로서는 절대적인 기준을 정할 수 있는 정보가 부족하기 때문에 K-RISS 점수의 분포 중에서 특정 기준치를 넘어가는 사안을 걸러낸다는 상대적인 측면의 모형이라고 볼 수 있다. 이에 향후 연구에서는 본 모형의 검증차원에서 실제로 위해 수준이 심각했던 상황들을 이용하여 본 연구에서 제시된 기준치와 비교하는 작업이 필요하다고 판단된다. 이는 기준치가 실제로 심각했던 상황에 대해서 잘 작동하는지 검증하는데 도움이 되며, 절대적인 기준치를 만드는데도 중요한 정보가 될 것이

다. 셋째, 시계열 모델을 통해 변화점을 탐지하는 것뿐만 아니라, 변화의 원인을 분석하여 모델에 적절히 반영하는 것이 중요하다. 원인까지 모델에 반영하는 것은 현재로서는 자료와 모형의 한계상 어려운 작업이라고 판단된다. 원인을 성격에 의해 구분하고 그것들이 일으키는 K-RISS 값의 분포에 대한 추가적인 분석을 거친 후 사안의 중요도 기준선에 대한 검토가 필요하다고 판단된다.

## 감사의 글

본 연구는 식품의약품안전처 “빅데이터 기반 안전관리 고도화를 위한 알고리즘 개발 연구(22063첨단약500)”의 연구개발비 지원에 의해 수행되었으며, 이에 감사드립니다.

## Conflict of Interest

No potential conflict of interest relevant to this article was reported.

## References

1. Bank of Korea. News Sentiment Index (NSI) released as experimental statistic. Available: <https://www.bok.or.kr/portal/bbs/B0000233/view.do?nttId=10070681&menuNo=200707> [accessed 20 April 2023].
2. Jeong YS, Choi HL, Yang DY, Kang EJ, Ko DK. Development of financial instability index and analysis of changes in financial instability factors. Sejong: Korea Institute for International Economic Policy; 2018 Dec. Report No.: 18-22.
3. Sah JY, Huh MY. A study on development of consumer safety indicators. Eumseong: Korea Consumer Agency; 2018 Dec. Report No.: 18-20.
4. Ministry of Food and Drug Safety. Announcement of 2020 Children's Dietary Safety Index survey results. Available: [https://www.mfds.go.kr/brd/m\\_99/view.do?seq=45953&srchFr=&srchTo=&srchWord=&srchTp=&itm\\_seq\\_1=0&itm\\_seq\\_2=0&multi\\_itm\\_seq=0&company\\_cd=&company\\_nm=&page=5](https://www.mfds.go.kr/brd/m_99/view.do?seq=45953&srchFr=&srchTo=&srchWord=&srchTp=&itm_seq_1=0&itm_seq_2=0&multi_itm_seq=0&company_cd=&company_nm=&page=5) [accessed 20 April 2023].
5. European Health Information Gateway. Environment and Health Information System (ENHIS). Available: <https://gateway.euro.who.int/en/datasets/enhis/> [accessed 29 March 2024].
6. U.S.-Mexico Border Field Office of the Pan American Health Organization. Environmental health indicators workshop U.S. - Mexico Border; 2002 Aug 28-30; Mc Allen, U.S.. Washington, DC: U.S.-Mexico Border Field Office of the Pan American Health Organization; 2002.
7. European Commission. DRMKC-INFORM. Available: <https://drmkc.jrc.ec.europa.eu/inform-index> [accessed 29 March 2024].
8. National Health Insurance Service, Health Insurance Review and Assessment Service. Drug assessment indicators of frequent occur-

ring diseases. In: National Health Insurance Service, Health Insurance Review and Assessment Service. editors. 2020 national health insurance statistical yearbook. Wonju: National Health Insurance Service, Health Insurance Review and Assessment Service; 2021. p.602-611.

9. Shewhart WA. Economic control of quality of manufactured product. New York: D. Van Nostrand Company; 1931.
10. Kim JG, Um SJ, Choi SW, Kim DN. Research results and trends analysis for EWMA control chart of manufacturing processes. Paper presented at: Korean Society of Safety Management Sciences 2013 Spring Conference; 2013 Apr 20; Anyang, Korea. p. 581-591.
11. Roberts SW. Control chart tests based on geometric moving averages. *Technometrics*. 1959; 1(3): 239-250.
12. Montgomery DC. Introduction to statistical quality control, 7th ed. Hoboken: Wiley; 2013.
13. Cho JH, Oh HS, Lee SJ, Jung SI, Lim T, Baek SS, et al. A comparative analysis on the efficiency of monitoring between EWMA and Shewhart chart in instrumental process with autocorrelation. *J Korean Soc Ind Syst Eng*. 2012; 35(4): 118-125.
14. Kim YB, Hong JS, Lie CH. Statistical design of Shewhart control chart with runs rules. *J Korean Soc Qual Manag*. 2008; 36(3): 34-44.
15. Andrews DWK. Tests for parameter instability and structural change with unknown change point. *Econometrica*. 1993; 61(4): 821-856.
16. Andrews DWK, Ploberger W. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*. 1994; 62(6): 1383-1414.
17. Bai J, Perron P. Estimating and testing linear models with multiple structural changes. *Econometrica*. 1998; 66(1): 47-78.
18. Bai J, Perron P. Critical values for multiple structural change tests. *Econom J*. 2003; 6(1): 72-78.
19. Liu J, Wu S, Zidek JV. On segmented multivariate regression. *Stat Sin*. 1997; 7(2): 497-525.
20. Lee SI. Comparison of control charts for individual observations. *Korean J Appl Stat*. 2022; 35(2): 203-215.

〈저자정보〉

진현정(교수), 허성윤(박사과정), 이현주(대표), 장보윤(대리)