



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](#)

한국형 입원환자분류체계(KDRG)
극단치 처리방법의 적정성에 관한 연구

연세대학교 보건대학원
보건통계학과 보건통계전공
김 진 휘

한국형 입원환자분류체계(KDRG)
극단치 처리방법의 적정성에 관한 연구

지도 남 정 모 교수

이 논문을 보건학 석사 학위 논문으로 제출함

2017년 6월 일

연세대학교 보건대학원
보건통계학과 보건통계전공

김 진 휘

김진휘의 보건학 석사 학위논문을 인준함

심사위원 남 정 모 
심사위원 박 소 희 
심사위원 김 윤 남 

연세대학교 보건대학원

2017년 6월 일

감사의 말씀

타지에서 업무와 학업을 병행하며 힘들기도 했지만 실무와 이론적 지식을 함께 채워가며 성장한다는 생각을 하고 즐기면서 지내다보니 어느덧 졸업을 하게 되었습니다.

통계적 이론 근거만으로는 해결할 수 없는 실무관련 통계 질문에도 경험적 노하우를 바탕으로 항상 명쾌한 해답과 끊임없는 조언 및 가르침을 주시고, 저에게 보건의료분야 통계에 대해서 또다른 새로운 시각을 열어주신 저의 스승님들이신 남정모, 박소희, 김윤남 교수님께 진심으로 감사드립니다.

건강보험심사평가원에서 연구를 할 수 있게 지원과 격려를 해주신 강경수, 공진선, 이충섭 실장님께 감사드립니다. 저를 의료분류체계 개발의 길로 인도해주시고, 따뜻한 관심과 아낌없는 가르침으로 항상 많은 도움을 주시는 김애련 부장님, 가장 가까운 곳에서 끊임없이 응원을 해주시는 차장님들과 나의 회사친구들인 의료분류체계실 과장들에게도 감사함을 전합니다.

무한한 사랑으로 저를 아껴주시는 아버지 김종웅님과 어머니 김행숙님께 사랑한다는 말씀을 전하고, 마지막으로 내 전부를 쥐도 아깝지 않을 유일한 사람 동생 김범휘에게 항상 고맙다는 말을 전합니다. 모두 감사합니다.

2017년 6월

김 진 휘 올림

차 례

국문요약	vi
I. 서 론	1
1. 연구배경 및 필요성	1
2. 연구 목적	4
II. 이론적 배경	5
1. DRG 개념 및 발전과정	5
2. KDRG 발전 및 적용분야	7
3. KDRG 구조 및 특성	11
4. 환자분류체계의 극단치 처리 방법	16
III. 연구방법	21
1. 연구모형	21
2. 연구자료 및 변수 정의	23
3. 분석 방법	25

IV. 연구결과	35
1. 분석자료의 일반적 특성	35
2. 분석대상 질병군의 진료비 분포 및 재원일수 관계 분석	44
3. KDRG 극단치 처리방법의 적정성 분석	56
4. KDRG 극단치 처리방법 대안 탐색 및 결과	66
V. 고찰	73
VI. 결 론	75
참고문헌	77
Abstract	81

표 차 례

표 1. KDRG의 MDC 체계	14
표 2. 국외 지불체계 DRG 극단치 절사 방법	19
표 3. 국외 지불체계 DRG 재원일수 극단치 상한값 계산 방법	20
표 4. 국외 지불체계 DRG 재원일수 극단치 하한값 계산 방법	20
표 5. 진료비와 재원일수의 극단치 유무 분할표	32
표 6. 분포 특성에 따른 극단치 처리 방법별 이상점	37
표 7. 분포 특성에 따른 극단치 처리 방법별 열외군 비율	37
표 8. MDC별 ADRG의 빈도	40
표 9. 분석대상 9개 질병군 ADRG	42
표 10. 9개 질병군의 빈도 및 왜도	43
표 11. 질병군별 진료비와 재원일수 상관분석	54
표 12. KDRG 자료특성에 따른 극단치 처리 방법	56
표 13. 질병군 F091 극단치 방법별 분할표 및 평가지수 비교 결과	57
표 14. 질병군 B050 극단치 방법별 분할표 및 평가지수 비교 결과	58
표 15. 질병군 N022 극단치 방법별 분할표 및 평가지수 비교 결과	59
표 16. 질병군 R020 극단치 방법별 분할표 및 평가지수 비교 결과	60
표 17. 질병군 E693 극단치 방법별 분할표 및 평가지수 비교 결과	61
표 18. 질병군 F704 극단치 방법별 분할표 및 평가지수 비교 결과	62
표 19. 질병군 O620 극단치 방법별 분할표 및 평가지수 비교 결과	63
표 20. 질병군 F770 극단치 방법별 분할표 및 평가지수 비교 결과	64

표 21. 질병군 L643 극단치 방법별 분할표 및 평가지수 비교 결과	65
표 22. 질병군별 진료비 극단치 처리 방법 적용시 열외군 비율	68
표 23. 질병군별 재원일수 극단치 처리 방법 적용시 열외군 비율	69
표 24. 열외군 비율이 낮은 극단치 처리 방법들 간의 일치도(ROC)	70

그 립 차 례

그림 1. KDRG 발진 과정	10
그림 2. KDRG 분류 과정	15
그림 3. KDRG 극단치 처리 방법 연구의 틀	22
그림 4. 분포에 따른 대표값들의 특징	29
그림 5. $N(0, \sigma^2)$ 인 경우 상자그림과 확률밀도함수	30
그림 6. 모집단의 분포가 $N(100000, 1000^2)$ 인 경우 히스토그램	38
그림 7. 모집단의 분포가 $LN(10, 1)$ 인 경우 히스토그램	38
그림 8. 질병군 F091 histogram & probability plot	45
그림 9. 질병군 B050 histogram & probability plot	46
그림 10. 질병군 N022 histogram & probability plot	47
그림 11. 질병군 R020 histogram & probability plot	48
그림 12. 질병군 E693 histogram & probability plot	49
그림 13. 질병군 F704 histogram & probability plot	50
그림 14. 질병군 O620 histogram & probability plot	51
그림 15. 질병군 F770 histogram & probability plot	52
그림 16. 질병군 L643 histogram & probability plot	53
그림 17. 질병군별 진료비와 재원일수 산점도	55
그림 18. 질병군별 진료비 극단치 처리 방법 적용시 열외군 비율	68
그림 19. 질병군별 재원일수 극단치 처리 방법 적용시 열외군 비율	69
그림 20. 열외군 비율이 낮은 극단치 처리 방법들 간의 일치도(ROC)	70

국 문 요 약

한국형 입원환자분류체계(KDRG) 극단치 처리방법의 적정성에 관한 연구

연구 배경

한국형 입원환자분류체계(Korean Diagnosis Related Group, 이하 KDRG)는 우리나라 입원환자의 진단 및 시술 등을 이용하여 자원소모 유사성과 임상적 유사성에 기초하여 환자들을 분류하는 체계이다. 현재 KDRG 극단치 처리 기준은 version 2.0(1991) 개발 당시 미국 Refined DRG 방법론을 기반으로 개발되었으며, 이후 두 차례의 전면 개정에도 동일한 방법론을 적용하였다. KDRG가 지속적으로 개정이 되면서 극단치 처리의 중요성을 인식하게 되었고, 극단치 처리 방법 기준 설정의 필요성이 제기되었다. 본 연구에서는 현재 KDRG에서 적용하고 있는 극단치 처리 방법의 문제점을 파악하고, 진료비 및 재원일수 자료의 특성에 상대적으로 더욱 적합한 극단치 처리 방법을 모색하고자 한다.

대상 및 방법

연구 자료는 2014년 1월 1일부터 12월 31일까지 건강보험심사평가원에 청구된 행위별 건강보험 입원진료 전체 7,691,707건을 KDRG version 4.0에 적용한 자료를 활용하였다. 분석대상은 질병군의 ADRG(Adjacent DRG) 수준에서 빈도와 분포를 고려하여 9개의 질병군을 대상으로 하였다. 진료비와 재원일수의 관계 및 분포를 확인한 후 현행 극단치 처리 방법과 함께 다양한 극단치 처리 방법들을 적용하여 상대비교를 하였다. 측정방법은 극단치 제거 비율 및 진료비와 재원일수 극단치의 일치도에 대한 통계량을 산출하여 상대 비교하였다.

연구결과

질병군의 진료비와 재원일수 자료의 특성을 알아보기 위해 기술통계량과 정규 적합도검정을 통해 분포를 확인하였다. 그 결과 진료비와 재원일수의 분포가 정규분포를 따르지 않으며, 오른쪽으로 꼬리가 긴 형태의 분포를 보였다. 그리고 분석대상 변수의 실효성을 확인하기 위해 상관분석을 실시한 결과 진료비와 재원일수 간 상관계수가 0.57~0.93으로 양의 상관관계를 보였다. 현행 방법을 포함한 극단치 처리 방법 5가지를 9개의 질병군에 적용하여 분석한 결과 현행 극단치 처리 방법을 적용할 경우 대부분의 질병군에서 극단치 비율이 5%를 초과하였으며, 일치도 측면에서도 다른 방법에 비해 상대적으로 일치도가 낮게 나타났다. 사분위수와 사분위수 범위를 활용한 현행 1단계 방법만을 적용한 경우 기타 방법들에 비해 상대적으로 극단치 비율이 낮았으며, 일치도도 높게 나타났다.

결론

KDRG 극단치 처리 방법의 적정성에 대한 본 연구를 통하여 극단치 처리전 데이터의 빈도와 분포 파악 및 극단치 제거로 인한 손실 정도를 파악하는 단계가 선행될 필요가 있다는 결론을 도출하였다. 분포의 특성에 맞는 극단치 처리 방법들 중 데이터 손실율이 적으며, 분석시 재원일수 또한 중요한 변수이므로 재원일수와의 극단치 일치도 측면을 고려해 볼 수 있다. KDRG 극단치 처리 방법 선정시 질병군별 빈도를 파악하고, 분포에 맞는 극단치 처리 방법을 적용하는 것이 적절할 것이며 KDRG의 활용 및 극단치 처리 목적에 따라 각기 다른 방안을 모색하는 것이 무엇보다 중요할 것이다.

I. 서 론

1. 연구배경 및 필요성

진단명기준환자군(Diagnosis Related Group, 이하 DRG)은 입원환자를 자원 소모 유사성과 임상적 유사성에 기초하여 분류하는 입원환자분류체계이다. DRG 분류체계에서는 모든 입원환자들이 주진단 및 기타진단, 수술명, 연령, 성별, 진료결과 등에 따라 진료내용이 유사한 질병군으로 분류되는데, 이 때 하나의 질병군을 DRG라고 부른다(한국보건산업진흥원, 2001). DRG는 입원환자를 대상으로 하는 포괄수가제도에서 지불단위로 사용될 뿐만 아니라 의료자원 이용감시, 병원경영, 의료의 질 관리 등 의료관리 분야에서 광범위하게 활용되고 있다(신영수 등, 1993).

현재 우리나라의 건강보험에서 사용되고 있는 한국형 입원환자분류체계(Korean DRG, 이하 KDRG)는 상병, 시술, 기능상태 등을 이용해서 입원환자를 의료 자원소모와 임상적 측면에서 유사한 그룹으로 분류하고 있다. 환자분류체계는 심사평가원 진료비 심사 및 평가업무시 환자구성상태(casemix) 보정 도구 및 각종 병원 간 진료비 수준 비교 자료로 이용되면서 활용범위가 지속적으로 확대되고 있다. 전문심사대상기관 선정 및 각종 심사·평가자료 지표 설정에도 이용하고 있으며, 상급종합병원 및 전문병원 지정에서 질병군 설정 기준으로도 이용되고 있다(대한의사협회, 2013).

DRG는 1977년 미국 Yale 대학에서 처음 개발되었다. 우리나라에 DRG가 소개된 것은 1986년 서울대학교 병원연구소 신영수 교수팀에 의해 한국형 진단명기준환자군이 개발되면서부터였다. 현재 미국의 공적의료보험을 관장하고 있는 보건복지부 산하 연방기관인 메디케어·메디케이드서비스센터(Center for

Medicare & Medicaid Services, 이하 CMS)의 전신인 보건의료재정청(Health Care Financing Administration, 이하 HCFA)이 1983년에 개발한 HCFA-DRG version 1.0을 근간으로 1986년에 우리나라의 KDRG version 1.0이 개발되었다.

이후 1991년에는 미국 Yale 대학의 Refined DRG 체계를 도입하여 기존의 KDRG에 중증도 분류체계를 접목시킨 KDRG version 2.0을 발표하였다. 2000년 8월부터의 개정작업에서는 관련 학회가 참여하여 KDRG의 임상적 타당성을 평가하고, 호주에서 개발된 AR-DRG의 원리에 따라 내과 시술 그룹의 신설, 중증도 분류방식의 변경 등 새로운 방식의 분류 원리를 도입 및 보완하여 관련 전산프로그램(KDRG Grouper)을 개발해 2003년 KDRG version 3.0을 발표하였다. KDRG version 3.0 발표 이후 전면 개정이 한동안 이루어지지 않자 최근의 임상현실을 반영한 개정의 필요성이 지속적으로 제기되어 2014년 1월부터 2년에 걸친 전면개정을 통해 2016년 KDRG version 4.0이 발표되었다. 전문의학회 및 심평원 내부 활용 부서 등에서 개선 건의한 내용 및 자체 청구자료 모니터링등을 토대로 실무검토, 임상전문가 자문회의를 거쳐 현재 KDRG version 4.1까지 개정이 되었다(건강보험심사평가원, 2017).

환자분류체계의 적정성을 모니터링하기 위해서는 분류체계의 기본 요소인 임상적·자원소모적 유사성을 판단하여야 하며, 이를 위해 자료원을 구축하고 자원소모 검토의 기본값인 진료비와 재원일수의 극단치를 제거하는 정제과정을 거치고 있다. 현재 사용되고 있는 KDRG 극단치 제거 방법은 두 단계로 구성되어 있다. 먼저 각 질병군별 제1사분위수(Q1)와 제3사분위수(Q3)를 구하여 사분위수범위(Interquartile range, 이하 IQR)를 이용하여 제거한다. 그리고 다시 기하평균(Geometric mean)과 표준편차(Standard deviation)를 이용하여 추가 제거하게 된다. 현행 방법은 KDRG version 2.0 개발 당시 미국 Refined DRG 방법론을 참조하여 만들어졌으며, 그간 version 3.0과 version 4.0 두 차례의 전면개정에서도 동일한 방법론을 적용하고 있다.

극단치는 어떤 지정된 그룹에 속하지 못하는 값으로 정상군의 상·하한 범위를 벗어난 자료들을 말하며, 분포에 따라 평균에 영향을 미쳐 신뢰도와 정확도를 떨어뜨리므로 극단치를 처리하는 것은 매우 중요하다. 극단치 정의 방법에 따라 열외군의 비율이 달라질 수 있으며, 동일한 방법에서도 질병군의 분포에 따라 차이가 있을 수 있다. 열외군의 비율은 5%이하가 이상적이며, 10%를 초과해서는 안된다(Palmer G 등, 2001).

질병군의 표본크기가 작은 경우 정규분포화가 어려우며, 표본의 크기가 큰 경우에도 10%이상 극단치 존재는 충분히 나타날 수 있다. 극단치 처리 방법 결정에 있어서 극단치 방법을 무리하게 적용하기보다는 KDRG 활용 목적에 따라 운영을 달리해야 한다(건강보험심사평가원, 2015).

현재까지 우리나라에서 KDRG 분류체계 질병군 극단치 처리에 관한 선행연구가 없었으며, KDRG 관련 연구 및 개정에서도 KDRG version 2.0(1991) 개발 당시 방법론을 현재까지 사용하여 연구하였기 때문에 제한점이 있다. 따라서 우리나라 건강보험 진료비와 재원일수 자료의 분포 및 특성을 통계적으로 고려한 연구가 필요하다.

이에 본 연구에서는 한국형 입원환자분류체계(KDRG)의 극단치 처리 방법을 개선하기 위해 가장 최신 자료인 KDRG version 4.0을 적용한 건강보험 행위별 입원청구 전수 자료를 활용하여 현행 극단치 처리 방법의 적정성을 검토하고자 한다. 또한 현행 KDRG 극단치 제거 방법과 기타 다른 극단치 처리 방법 적용한 것을 비교 분석하여 우리나라 자료에 적합한 극단치 처리 방법을 모색하고자 한다.

2. 연구 목적

본 연구의 목적은 한국형 입원환자분류체계(KDRG)의 극단치 처리 방법을 개선하기 위해 KDRG version 4.0에 대해 현행 극단치 처리 방법을 적용하여 적정성을 검토하고, 그 외 다양한 극단치 처리 방법과 비교하여 적합한 극단치 처리 방법을 모색하는 것이다.

세부적인 연구목적은 다음과 같다.

첫째, KDRG 분류체계의 기본 질병군 단계인 ADRG 수준에서 질병군별 빈도와 왜도를 포함한 기술통계량과 정규적합성 검정을 통해 분포를 확인하여 건강보험 청구 자료의 특성을 파악한다.

둘째, 질병군별 진료비와 재원일수의 상관분석을 통해 상관정도를 확인하여 분석변수 대상의 실효성을 확인한다.

셋째, 빈도 및 분포를 고려한 다양한 질병군을 대상으로 하여 진료비와 재원일수에 현행 방법과 이 외의 방법들을 적용한다. 극단치의 제거 비율과 함께 극단치 사례의 일치도를 통계적으로 분석하여 결과를 상대 비교함으로써 적정성을 평가한다.

II. 이론적 배경

1. DRG 개념 및 발전과정

진단명기준환자군(Diagnosis Related Group, 이하 DRG)은 입원환자를 임상적·자원소모적 유사성에 기초하여 분류하는 입원환자분류체계이다. DRG의 개발은 병원의 최종 산출물(final product)을 정량화하여 병원 간 비교나 원가분석에 사용하고자 하는 노력의 일환이었다(한국보건산업진흥원, 2001). 병원의 경우 노동력, 물자, 시설과 기구들을 재료로 사용하여 간호, 제약, 그리고 검사 등의 서비스를 생산한다고 볼 수 있고 각 환자들은 의사의 지시에 따라 이러한 서비스의 특정한 묶음을 제공받는다고 할 수 있다(강길원, 2003). 환자들의 특성에 따라 의약품과 검사 서비스 등이 다양해지고 재원일수 및 진료비도 달라진다. 환자가 필요로 하는 서비스의 종류와 양을 결정하는 질병명과 치료방법 등 여러 요인에 따라 분류하면 임상적으로 유사한 묶음의 서비스를 필요로 하는 환자군을 정의할 수 있으며, 이를 DRG라고 부른다.

1965년 미국의 Medicare 출범 이후 보험자로부터 의료비를 상환 받는 조건으로 미국의 병원들은 이용도 조사(utilization review)나 질 보장(quality assurance)에 관한 자료의 제출이 의무화되었다. 이로 인해 병원에서의 의료서비스 활동을 어떻게 측정할 것인가 또는 일반산업부문에서 사용되고 있는 비용과 품질에 관한 경영관리 기법을 의료부문에 그대로 적용할 수 있을 것인가에 관심을 가지게 되었다(가와부치 코우이치, 2000). 1969년 미국의 Yale 대학에서는 병원의 활동을 측정하기 위한 연구에 착수하며 DRG의 설계 및 개발이 시작되었다. Yale 대학 연구팀은 병원의 중간 산물(intermediate products)에 해당하는 원재료(노동력, 소모품, 장비 등)를 표준화된 산출물(식사, 소독된

리넨, 임상병리검사, 투약 등)로 전환하였다. 그리고 병원의 주된 업무인 질병이나 장애를 가진 입원환자들을 평가하고 치료하며 의사나 전문가들의 지도하에 병원은 해당 질병의 진단과 치료에 필요할 것이라고 판단되는 일련의 제품과 서비스를 제공하게 되는데 Yale 대학 연구팀은 병원의 산출물로서 묶음에 주목하였다.

Yale 대학의 접근 방법은 개개의 서비스보다는 특정 질병을 가진 환자에게 제공되는 물품이나 서비스의 묶음으로 정의되는 병원의 최종산출물에 초점을 두었다는 점에서 전통적인 접근 방법과는 달랐다. 이러한 초점의 변화로 서비스 생산의 효율성(efficiency)뿐만 아니라, 서비스 사용의 효과성(effectiveness)에 대한 고려가 가능해졌다(강길원, 2003). DRG의 목적은 병원의 환자구성을 자원소모와 연결시키고자 하는 것으로, 환자의 유형을 결정하고 자원소모량을 예측하는 조작적인 방법 개발이 필요하게 되었다. 임상적으로 유사하고 자원소모가 유사한 환자그룹을 개발함으로써, 환자를 그룹화하면 병원의 환자구성을 예측하고 측정할 수 있게 된다. 따라서 임상적인 측면과 자원소모량 두 가지 측면에서 유사한 환자그룹을 구성하는 환자분류체계의 한 부분으로 DRG가 개발되었다(OTA, 1983).

2. KDRG 발전 및 적용분야

우리나라에 DRG가 소개된 것은 1986년 서울대학교 병원연구소에 의해 한국형 진단명기준환자군(KDRG)이 개발되면서부터였다. 당시 KDRG version 1.0은 미국 Medicare에서 사용하는 Medicare DRG를 기초로 만들어졌다. 이후 1991년에는 Yale 대학의 RDRG 체계를 도입하여, 기존 KDRG에 중증도 분류체계를 접목시킨 KDRG version 2.0이 발표되었다. 그러나 KDRG를 개발할 당시 포괄수가제도에 대한 거부감으로 인해 의료계의 참여가 거의 이루어지지 못했고, 이용할 수 있는 데이터베이스도 한정되어 있었기 때문에 미국 DRG의 규정을 대부분 수용하여 KDRG가 만들어졌다. 하지만 미국과 우리나라는 질병구조나 진료행태가 다르기 때문에 미국 DRG를 그대로 사용하는 것은 부적합할 수 있다. DRG를 도입하여 사용하고 있는 호주나 유럽의 국가들도 미국 DRG를 그대로 도입하여 사용하기보다는 자국의 실정에 맞게 일부 수정하거나 자체 개발하여 사용하고 있다는 점을 고려하면 우리나라의 진료행태나 필요에 맞게 미국 DRG를 수정하여 한국형 DRG를 만드는 과정이 필요하였다. 이러한 필요성 때문에 KDRG 개정 작업이 2000년 8월부터 2002년 12월까지 2년 반에 걸쳐 진행되었으며, 기존과는 달리 관련 의학회가 참여하여 KDRG의 임상적 타당성을 평가하였다. 이후 개정된 내용을 일부 보완하고 관련 전산프로그램(KDRG grouper)을 개발하여 2003년 12월에 KDRG version 3.0이 발표되었다.

KDRG version 3.0 발표 이후, 분류체계에 대한 공개적인 의견 수렴 과정을 거쳐서 일부 내용을 수정하고, 새롭게 신설된 수가코드와 100/100 본인부담 행위코드를 분류체계에 반영하여 KDRG version 3.1을 2005년 2월에 발표하였다. 그리고 2005년 2월 이후부터 2008년 5월까지 삭제 및 신설된 수가코드와 2008년 1월부터 적용되는 한국표준질병·사인분류 5차 개정을 DRG 분류에 반영한 KDRG version 3.2 개정에 이어 의료환경의 변화와 분류체계와 관련하여

의료계 및 관련 기관 등에서 건의한 내용들을 폭 넓게 반영한 KDRG version 3.3이 2009년부터 적용되었다. 2011년 1월부터 적용된 KDRG version 3.4에서는 한국표준질병·사인분류 6차 개정 내용을 반영하였으며 진단코드가 기존의 5자리에서 6자리까지 세분화되었다. 이후 KDRG 활용 부서 및 관련 기관 등에서 건의한 내용 등을 반영하여 임상학회의 의견수렴을 거쳐 개정한 KDRG version 3.5가 2014년 1월부터 적용되었다.

KDRG version 3.0 개발(2003) 이후 전면 개정이 이루어지지 않아 최근의 임상현실 반영을 위한 개정의 필요성이 지속적으로 제기되었다. 이러한 필요성 때문에 2014년 1월부터 2015년 12월까지 2년에 걸쳐 KDRG version 4.0 전면 개정 작업이 진행되었으며 기존과는 달리 개정 시작단계부터 종료 시까지 전문의학회 임상전문가 200여명이 참여하여 임상적 타당성을 평가하였고 통계전문가 자문도 수차례 진행하였다.

KDRG version 4.0 개정은 단계적으로 추진되었다. 2014년에는 의료계가 수행한 임상적 타당성 연구결과를 토대로 질병군(ADRG) 구성에 대한 검토가 이루어졌고, 2015년에는 분류기준 및 분류구조와 중증도 등에 대한 개선 검토가 이루어졌다. 이후 전산프로그램(KDRG grouper)을 개발하여 KDRG version 4.0을 2015년 12월에 발표하여 2016년 1월부터 적용하고 있다. 최근 전문의학회 및 심평원 내부 활용부서 등에서 개선 건의한 내용 및 청구자료 모니터링한 내용을 토대로 실무검토, 임상전문가 자문 및 회의를 거쳐 KDRG version 4.1까지 개정이 완료된 상태이다(건강보험심사평가원, 2017).

DRG는 포괄수가제도의 지불단위로서 사용될 뿐만 아니라 의료자원 이용감시, 병원경영, 의료의 질관리 등에 사용되고 있는 관리도구이다. 현재 KDRG는 질병군별 포괄수가제(7개 질병군 및 실포괄 지불)에서 지불단위로 사용되고 있으며, 상급종합병원 지정 및 전문병원 인정 기준에 의료기관 기능 평가 도구로 사용되고 있다. 또한 의료기관 간 진료비나 질적 수준을 비교하기 위

해서는 비교 대상이 되는 의료기관의 환자구성(casemix)이 동일해야 하는데
의료기관마다 환자구성이 다르기 때문에 이를 보정하기 위한 도구가 필요하
다. 진료비용, 재원일수, 사망률, 기타 질 지표 등을 병원 간 비교시 환자구성
보정 도구로서 환자분류체계가 가장 널리 사용되고 있다.

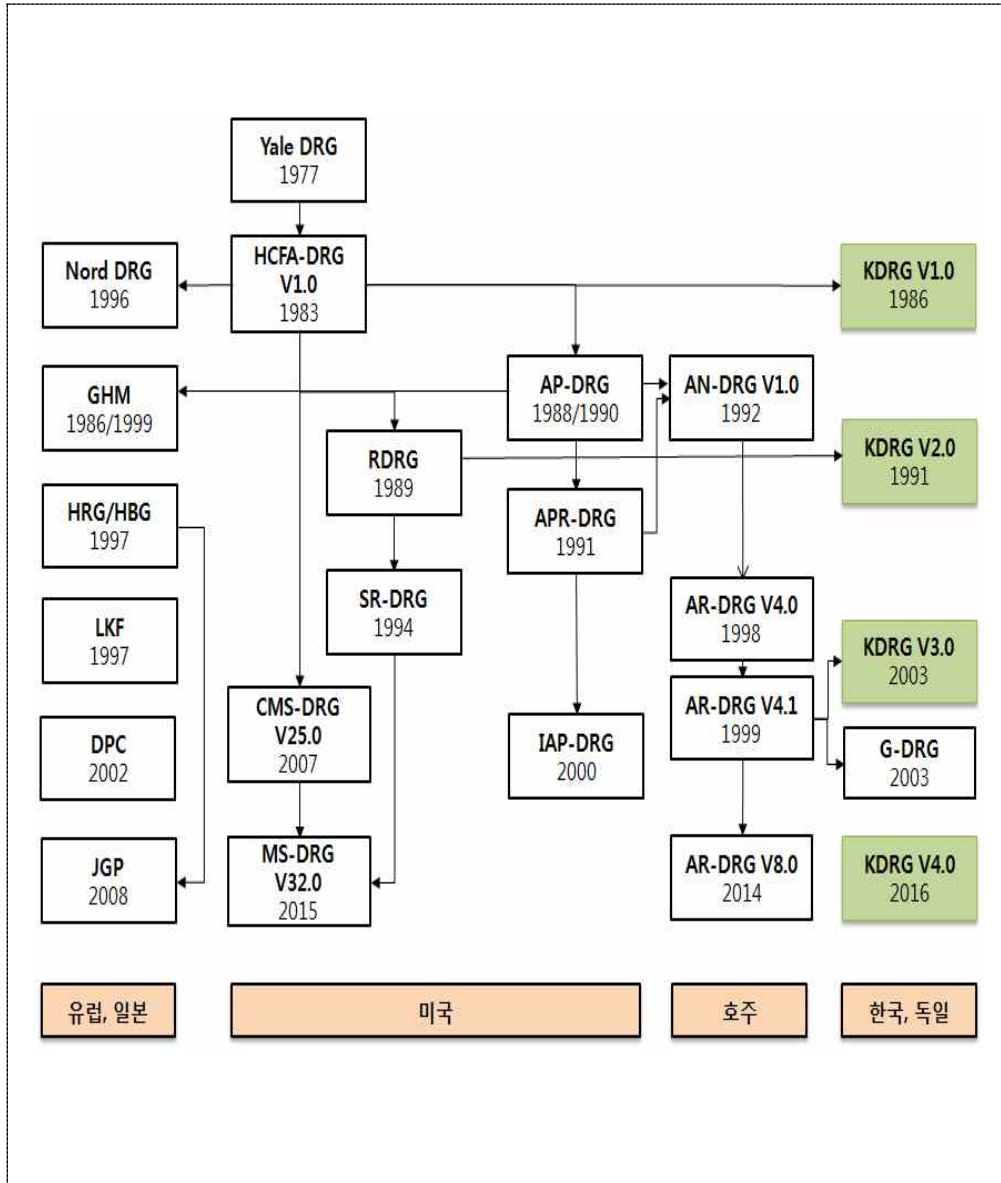


그림 1. KDRG 발전 과정.

3. KDRG 구조 및 특성

DRG 분류체계의 개발에서 DRG를 정의하기 위해 사용되는 변수는 병원에서 일상적으로 수집되는 전산화된 정보여야 하며, DRG의 개수는 관리할 수 있는 범위 내에 있어야 한다. 또한 각 DRG는 자원소모가 유사한 환자를 포함하며, 임상적 관점에서 유사한 환자를 포함해야 하는 4가지 원칙을 가지고 있다(한국보건산업진흥원, 2001). 현재는 정보통신기술의 발전과 병원의 전자의무기록이 도입되고 진료비 관리부분도 모두 전산화가 되어있다. 임상적 타당성을 확인하기 위해 보험 청구자료 이외의 병원자료를 확보하여 분석할 필요성이 있으며, DRG 관리 개수의 제한도 정보시스템의 발달에 따라 함께 고려를 해야 하므로 초창기 분류 원칙 기준을 현실에 맞게 일부 수정하여 KDRG의 임상적 타당성 연구를 진행해야 한다(대한의사협회, 2013).

KDRG의 분류과정은 크게 4단계 과정(그림 2)으로 볼 수 있는데 1단계는 주진단에 따라 입원환자를 26개 주진단범주(Major Diagnostic Category, 이하 MDC) 중 하나로 분류를 한다(표 1). 2단계에서는 환자가 수술을 받았는지 여부에 따라 외과계와 내과계 질병군으로 구분이 된다. 그리고 3단계에서 외과계는 수술 유형, 내과계는 주진단(조기사망제외)에 따라 질병군 세분화가 된다. 이 단계까지의 분류를 ADRG(Adjacent DRG)라고 하며, ADRG는 기본 DRG(base DRG)의 개념이다. 마지막으로 4단계는 연령구분과 기타진단을 이용한 중증도분류로 최종 DRG를 결정한다. 이 단계까지의 분류를 RDRG(Refined DRG)라고 한다. RDRG는 중증도가 반영된 final DRG의 개념이다.

KDRG version4.0의 분류번호 구조는 알파벳과 숫자를 혼용하여 6자리로 구성되어 있다. 첫번째 자리 알파벳은 주진단 범주를 나타내는 MDC이고, Error DRG의 경우 알파벳 대신 숫자 9로 표기가 된다. 두번째부터 세번째 자리 숫자는 ADRG 분류를 나타낸다. 01~49는 외과계 그룹, 50~59는 내과적 시술 그

룹, 60~99는 내과계 그룹을 뜻한다. 네번째 자리 숫자는 ADRG의 소분류를 나타내며 0인 경우 질병군 세분류가 없는 경우를 뜻하고, 1~9와 A~Z인 경우는 시술코드나 진단코드를 임상적·진료비 측면이 유사하도록 세분류가 된 것을 의미한다. 알파벳 O와 I는 숫자와의 혼선을 피하기 위해 제외된다. 다섯번째 자리 숫자는 연령구분의 숫자로 연령구분이 없는 경우 0으로 표기되고, 있는 경우는 1~3으로 ADRG별 연령구분에 따라 차례대로 부여된다. 그리고 마지막 여섯번째 자리 숫자는 중증도를 구분하는 자리로 중증도 미동반의 경우 0으로 표기되며, 중증도를 동반하는 경우 ADRG별 중증도 구분에 따라 차례대로 부여가 된다.

KDRG에서 중증도 구분은 CCLs와 PCCL이라는 결정 방법을 가지고 있다. CCLs(Complication and comorbidity levels)는 모든 기타 진단에 부여된 중증도 가중치(severity weight)를 말하며, 신생아나 외과환자의 경우 0~4까지의 점수를 가지고, 내과환자의 경우 0~3까지의 점수를 가진다.

- CCL = 0 : - CC가 아닐때
 - 환자가 배정된 ADRG의 정의에 포함된 코드일 때
 - 주진단과 밀접하게 연관된 CC일 때
 - 해당 레코드의 다른 코드와 정확하게 일치할 때
- CCL = 1 : minor CC
- CCL = 2 : moderate CC
- CCL = 3 : severe CC
- CCL = 4 : catastrophic CC

PCCL(Patient clinical complexity level)은 환자가 가진 합병증이나 동반상병 (Complication and comorbidity, 이하 CC)의 누적 효과를 나타내는 척도이다. 일반적으로 중요 합병증 혹은 동반상병을 CC로 표기한다. PCCL값은 유사한 질병이 두번 반영되지 않도록 고안되었으며 매 입원진료건마다 PCCL 공식에 의해 계산되어진다(강길원, 2003).

PCCL = 0 : no CC effect
 PCCL = 1 : minor CC
 PCCL = 2 : moderate CC
 PCCL = 3 : severe CC
 PCCL = 4 : catastrophic CC

PCCL = 0 if there is no SDx
 = 4 if $x > 4$
 = x otherwise

$$x = \text{round} \left[\frac{\ln(1 + \sum_{i=k} CCL(i) \times \exp(-\alpha \times (i-k)))}{\ln(3/\alpha)/4} \right]$$

$\alpha = 0.4$

$k = 1$ for *ADRG* P01 – P67, $k = 2$ for all other *ADRGs*

표 1. KDRG의 MDC 체계

MDC		Description
MDC 00	A	PreMDC
MDC 01	B	신경계의 질환 및 장애
MDC 02	C	눈의 질환 및 장애
MDC 03	D	이비인후, 구강의 질환 및 장애
MDC 04	E	호흡기계의 질환 및 장애
MDC 05	F	순환기계의 질환 및 장애
MDC 06	G	소화기계의 질환 및 장애
MDC 07	H	간담도계 및 췌장의 질환 및 장애
MDC 08	I	근골격계 및 결합조직의 질환 및 장애
MDC 09	J	피부, 피하조직, 유방의 질환 및 장애
MDC 10	K	내분비, 영양, 대사성 질환 및 장애
MDC 11	L	신장 및 비뇨기계의 질환 및 장애
MDC 12	M	남성 생식기계의 질환 및 장애
MDC 13	N	여성 생식기계의 질환 및 장애
MDC 14	O	임신, 출산, 산욕
MDC 15	P	신생아
MDC 16	Q	혈구 및 조혈기관의 질환과 면역장애
MDC 17	R	신생물 질환(혈액 및 고형암)
MDC 18-1	S	전염성 및 기생충 질환(HIV 감염)
MDC 18-2	T	전염성 및 기생충 질환
MDC 19	U	정신 질환 및 장애
MDC 20	V	알코올/약물남용 및 알코올/약물로 인한 기질적 정신 장애
MDC 21-1	W	다발성 외상
MDC 21-2	X	손상, 중독, 약물의 독성작용
MDC 22	Y	화상
MDC 23	Z	건강 상태 및 보건의료서비스 이용에 영향을 미치는 요소

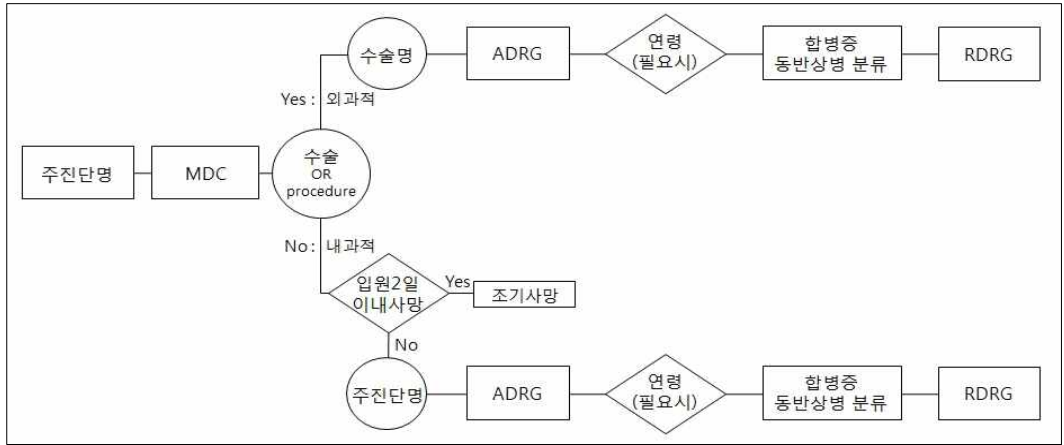


그림 2. KDRG 분류 과정.

4. 환자분류체계의 극단치 처리 방법

환자분류체계는 질병군별 자원소모 유사성의 동질성을 지향하며, 자원소모 검토의 기본값으로 평균진료비와 평균재원일수를 산출하게 된다. 질병군 그룹 내에서 대부분 자원소모가 유사한 환자들로 구성되어 있지만, 일부 극단적인 값을 가진 환자도 함께 포함이 된다. 특히 진료비 자료의 경우 극단치가 매우 큰 값을 가지는 경우가 많아 진료비가 과대평가되고 이러한 특성에 따라 우측으로 꼬리가 긴(right skewness) 분포의 형태를 나타낸다. 극단값이 질병군의 평균 진료비 및 재원일수에 영향을 미칠 수 있으므로 평균값에 대한 과대평가를 방지하기 위해 극단치를 처리하는 정제 과정을 거치고 있다.

극단치의 처리 방법에 대해 여러 가지 방법이 제안되고 있으나 자료의 분포나 특성에 따라 사용되는 방법이 상이하다. 일반적으로 모수적·비모수적 방법 2가지로 구분하여 많이 사용되고 있다. 모수적 방법은 자료가 정규 분포를 따른다는 가정으로 자료의 평균과 표준편차를 이용하여 이상점을 정의하는 것이다. 로그노말분포를 가정하고 자연로그 변환 후 이상점을 찾은 뒤 지수변환을 통해 원래 값으로 반환하거나 기하평균값을 사용하는 방법도 제시되고 있다. 그리고 비모수적 방법은 분포를 가정하지 않고 값의 순위를 이용한다. 제1사분위수와 제3사분위수를 기준으로 사분위수범위(IQR)를 이용하는 것이 대표적이다. 상수 값은 일반적으로 1.5~3.0 사이의 값을 사용하며 3을 대입하였을 때 초과하는 값들을 극단치라고 정의하기도 한다. 이상치의 비율은 5%이하가 이상적이며, 10%를 초과해서는 안된다(Palmer G 등, 2001).

KDRG 극단치 처리 방법은 KDRG version 2.0(1991) 개발 당시 열외군 개념을 동일 집단내에 다른 특성을 가진 값이 포함되어 있다는 가정에서 출발하여, 극단치를 제거하는 개념으로 만들어졌으며 현재까지 동일한 방법을 적용하고 있다. 열외군은 먼저 각 질병군별로 제1사분위수(Q1)와 제3사분위수(Q3)

를 구하여 아래(식1)에서 구해진 상하한의 범위를 벗어나는 자료들을 제외한 후, 1차 제거된 데이터를 가지고 다시 기하평균과 표준편차를 구하여 (식2)에서 구해진 상·하한의 범위를 벗어난 진료건을 제외한다.

$$(식1) \quad Q1 - 3 \times (Q3 - Q1), \quad Q3 + 3 \times (Q3 - Q1)$$

$$(식2) \quad \text{기하평균} - 3 \times \text{표준편차}, \quad \text{기하평균} + 3 \times \text{표준편차}$$

1단계에서는 극단치의 영향을 받지 않는 사분위수와 사분위수 범위를 이용하였으며, 극단치의 최소제거를 위해 상수를 3으로 하여 사분위수 범위의 3배를 적용하였다. 2단계에서는 극단치의 영향을 비교적 덜 받는 기하평균을 location measurement로 활용하였으며 표준편차는 displacement measurement 개념으로 상수를 3으로 하여 3배를 적용함으로써 극단치를 최소한으로 제거하였다(건강보험심사평가원, 2015).

한국형 환자분류체계에서도 입원(KDRG)·외래(KOPG)·한의입원(KDRG-KM)·한외래(KOPG-KM)의 극단치 처리 방법이 다르다. KOPG의 경우 여러 극단치 처리 방법론을 대상으로 질병군별 데이터 손실율과 분산의 상대적 감소 정도를 나타내는 RIV(Reduction in Variance)을 확인하여 극단치 처리 방법(식2)을 적용하고 있으며, KDRG-KM·KOPG-KM의 경우 전문가 자문을 통해 자료 분포를 고려한 극단치 처리 방법(식1)을 적용하고 있다.

환자분류체계에서는 질병군별 동질한 그룹 구성을 위해 극단치를 처리한다 면, 지불체계에서는 재정적 요소와 위험분산, 수가 모형등을 고려하여 극단치를 처리한다. 지불체계에서 극단치와 고가 진료에 대한 지불 정확성을 높이기 위해 DRG를 지속적으로 개선했음에도 불구하고, 모든 국가의 DRG 그룹들은 같은 DRG내에서 대부분의 환자보다 더 많은 자원을 소모하는 환자들을 포함하고 있다. 이러한 고가 극단치 사례들은 일반적으로 전체 병원 비용의 상당

한 부분을 차지하고, 결과적으로 DRG 사례들의 평균 비용에 상당한 영향을 준다(Cots F 등, 2003).

극단치 사례를 따로 분리해서 지불하지 않을 경우 병원은 고비용 사례를 피하려 하거나, 환자들을 부적절하게 조기 퇴원시키려는 유인을 갖게 될 것이다. 결과적으로 대부분의 국가들은 극단치 사례들을 식별하고 극단치 환자들의 추가 비용에 대해 추가적으로 병원지불을 하는 방법을 개발해왔다. 한편, 극단치 제거시 국가별로 자원소모 분석변수가 다양하다. 많은 유럽국가에서 표준화되고 비교 가능한 양질의 원가 정보는 제한되어 있다. 환자 수준에서 양질의 원가 정보가 없는 경우 일부 국가에서는 환자의 재원기간을 기준으로 한 자체 DRG 시스템을 개발하였고, DRG식 환자분류체계에서 환자 분류를 위해 재원일수는 가장 많이 사용되는 자원소모 분석변수이다(Busse R 등, 2012).

표2에서 볼 수 있듯이 대부분의 국가들은 재원일수 한계치를 근거로 극단치 사례를 정의한다. 노르딕 국가들(에스토니아, 핀란드, 스웨덴)은 예외적으로 비용을 이용해 극단치를 정의한다. 그러나 네덜란드를 제외한 모든 국가들이 극단치를 정의할지라도 극단치를 결정하는 방법은 국가마다 다르다(Busse R 등, 2012).

DRG 시스템의 목적과 데이터 특성에 기반하여 극단치 처리 방법을 선택해야 하기 때문에 최상의 극단치 처리 방법은 없다. 독일, 프랑스와 같이 일부 국가에서는 단기 입원으로 인한 낮은 비용을 설명하기 위해 LOS의 극단치 하한 기준(표4)을 결정한다. 반면 다른 국가에서는 단기 입원으로 인한 인센티브를 창출하기 위해 명시적으로 극단치 하한 기준을 사용하지 않는다(Schreyogg J 등, 2006).

표 2. 국외 지불체계 DRG 극단치 절사 방법

Country	trimming method
Austria	LOS (interquartile)
England	LOS (interquartile)
Estonia	Cost (parametric)
Finland	Cost (parametric)
France	LOS (interquartile)
Germany	LOS (parametric)
Ireland	LOS (parametric)
Netherlands	-
Poland	LOS (interquartile)
Portugal	LOS (interquartile)
Spain(Catalonia)	LOS (interquartile)
Sweden	Cost/LOS (parametric)

Source : Diagnosis-Related Groups in Europe, 2011.

표 3. 국외 지불체계 DRG 재원일수 극단치 상한값 계산 방법

Country	Method Used
Germany	$LOS\text{-threshold}_i = round[\min[mean_i + 2 \cdot SD_i; mean_i + 17]]$
England/ Denmark	$LOS\text{-threshold}_i = Q75_i + (Q75_i - Q25_i) \cdot 1.5$
Italy	$LOS\text{-threshold}_i = (\sqrt[3]{Q75_i} + (\sqrt[3]{Q75_i} - \sqrt[3]{Q25_i}) \cdot 1.0)^3$
France	$LOS\text{-threshold}_i$ (if $mean_i > 8$ days) $= \min[15; \exp^{mean[\log(los_i)] + SD[\log(los_i)] + Q95_i - median[\log(los_i)]}]$ $LOS\text{-threshold}_i$ (else) $= \min[2.5 \cdot mean_i; \exp^{mean[\log(los_i)] + SD[\log(los_i)] + Q95_i - median[\log(los_i)]}]$
Spain	$LOS\text{-threshold}_i = median_i + 3 \cdot SD_i$ or, $LOS\text{-threshold}_i = Q75_i + (Q75_i - Q25_i) \cdot 1.5$ or, $LOS\text{-threshold}_i = Q75_i + (Q75_i - Q25_i) \cdot 2.0$

i refers to the standard deviation(SD), mean, median or LOS-threshold of the i th DRG.

Source : Methods to determine reimbursement rates for diagnosis related groups (DRG):
a comparison of nine European countries, 2006.

표 4. 국외 지불체계 DRG 재원일수 극단치 하한값 계산 방법

Country	Method Used
Germany	$LOS\text{-threshold}_i = round[\min[2; Mean_i/3]]$
France	$LOS\text{-threshold}_i = \min[(Mean_i/2.5); 7; (mod - 1)]$ (only if $mean_i \geq 8$, else no lower LOS-threshold)

i refers to the mean, median or LOS-threshold of the i th DRG.

Source : Methods to determine reimbursement rates for diagnosis related groups (DRG):
a comparison of nine European countries, 2006.

Ⅲ. 연구방법

1. 연구모형

이 연구의 목적은 한국형 입원환자분류체계(KDRG)의 극단치 처리 방법을 전반적으로 검토하여 현행 방법의 문제점을 파악하고, 우리나라 입원자료의 특성에 적합한 새로운 극단치 처리 방법을 탐색하여 KDRG 개정의 합리성 및 신뢰성을 제고하는 것이다. 이를 위해 국내외 환자분류체계 극단치 처리 방법에 대한 문헌고찰을 선행하고 실제 우리나라 입원자료원을 이용하여 질병군별 진료비와 재원일수의 분포 및 특성을 고려한 방법론들을 적용하였다. 이후 통계분석을 실시하여 결과를 상대 비교함으로써 우리나라 자료의 특성에 적합한 극단치 처리 방법을 탐색하였다. 연구의 틀은 다음과 같다(그림 3).

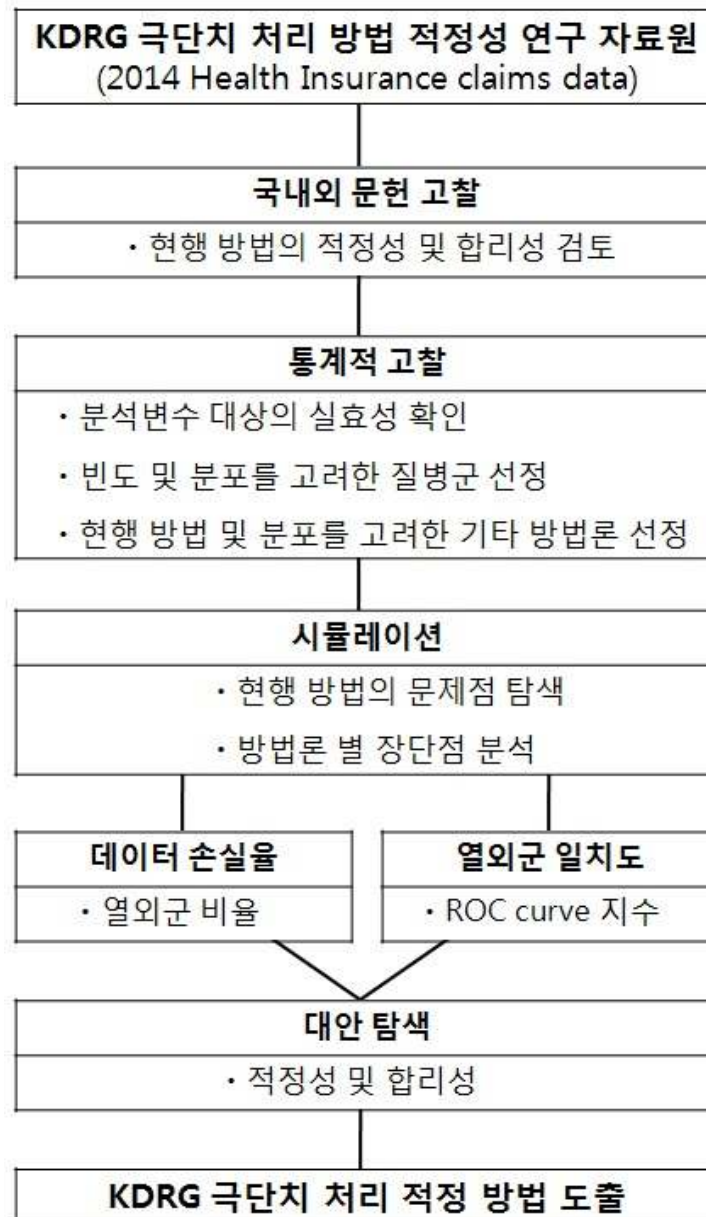


그림 3. KDRG 극단치 처리 방법 연구의 틀.

2. 연구자료 및 변수 정의

가. 연구자료

극단치 처리 방법 측정을 위해 2014년 1월부터 12월까지(심사차수 기준 2015 3월 31일까지) 건강보험심사평가원에 청구된 건강보험 행위별 입원청구 자료 7,691,707건을 KDRG version 4.0에 적용한 자료를 활용하였다. 월별 분리 청구 건에 대해서는 환자별로 합산하여 에피소드별로 구축하였다. 추가청구건의 경우 누락된 진료내역만 추가로 청구하는 건으로 원청구와 합산되지 않을 경우 의미가 없으므로 제외하였으며, 보완청구의 경우는 심사불능 처리되었던 원청구를 보완하여 재청구하는 것으로 원청구와 동일한 것으로 판단하여 대상에 포함하였다. 입원료 최저금액(보건의료원/낮병동) 미만인 건을 제외하였고, 주진단 범주 분류에서 진료비 변이가 큰 시술로 이루어진 PreMDC와 Error DRG는 범주 유형 특성상 분석에서 제외하였다. 요양기관 종별로는 의원, 병원, 종합병원, 상급종합병원을 분석대상에 포함하였다.

분석 대상은 질병군의 빈도와 분포를 고려하여 전체 질병군 중 빈도가 25 percentile 이하인 질병군, 25percentile을 초과하고 75percentile 미만 질병군, 75percentile 이상 질병군으로 그룹을 나누었다. 이후 질병군 빈도 기준으로 나눈 세그룹 내에서 질병군 각 진료비의 왜도값이 25percentile, median, 75percentile에 위치하며, 극단치 처리 방법들을 모두 적용하였을 때 진료비와 재원일수에서 극단치가 발생하는 질병군을 대상으로 하여 추출된 9개의 질병군이 분석 대상으로 선정되었다.

나. 변수 정의

질병군 별 진료비와 재원일수에 극단치 처리 방법들을 적용하여 열외군과 비열외군에 대해 상대적으로 가장 큰 일치를 나타내는 방법이 가장 만족스러운 극단치 처리 방법이라는 가설을 두고 있다(Cots F 등, 2003).

분석 변수는 KDRG version 4.0 질병군 변수와 진료비 및 재원일수로 하였다. 진료비의 경우 자원소모량을 정확하게 반영하기 위해 입원료 및 종별 가산율을 상급종합병원으로 표준화시킨 표준화된 총진료비를 사용하였다.

3. 분석 방법

분석자료(2014년 건강보험 입원청구자료)의 일반적 특성을 파악하기 위하여 기술통계량을 산출하고, 질병군 분포 특성에 적합한 방법론을 적용하고자 정규적합성 검정을 통하여 질병군의 분포를 확인하였다. 진료비와 재원일수간의 관계를 알아보기 위해 상관분석을 실시하였으며, 일치성을 알아보기 위해 진료비와 재원일수의 극단치 유무를 이용한 분할표(contingency table)를 생성하여 accuracy, Kappa coefficient, sensitivity, specificity, ROC curve 분석을 하였다. 또한 정상군 대비 극단값의 정도를 알아보기 위해 진료비와 재원일수에서 극단치 제거비율을 확인하였다. 모든 통계분석은 SAS version 9.4(SAS Institute Inc., Cary, NC, USA.)를 사용하였다.

가. KDRG의 현행 극단치 처리 방법

현재 사용되고 있는 KDRG 극단치 제거방법은 KDRG version2.0(1991) 개발 당시 미국 Refined DRG 방법론을 참조하여 만들어졌으며, KDRG version3.0 개발 당시 열외군 정의를 토대로 KDRG version4.0 개정에도 동일한 방법론을 적용하고 있다. 먼저 각 ADRG별로 제1사분위수(Q1)와 제3사분위수(Q3)를 구하여 아래 (식1)에서 구해진 상하한의 범위를 벗어나는 자료들을 제외한 후, 다시 기하평균과 표준편차를 구한다. 여기서 구해진 값을 이용하여 (식2)의 상하한의 범위를 벗어난 건을 제외한다(강길원 외, 2003). KDRG version2.0 개발 당시 outlier 개념은 동일집단 내에 '다른 특성'이 들어있다는 가정에서 출발하여, extreme observation을 제거하는 개념이다. 그러므로 extreme observation

의 최소 제거를 위해 상수값을 3으로 설정하였다(건강보험심사평가원, 2015).

(식1) $Q1 - 3 \times (Q3 - Q1)$, $Q3 + 3 \times (Q3 - Q1)$

(식2) (기하평균 $- 3 \times$ 표준편차), (기하평균 $+ 3 \times$ 표준편차)

나. 자료특성에 따른 극단치 처리 방법 탐색

환자분류체계 관련 분석의 경우 종속변수로서 원가자료(cost)를 이용한 분석이 최선이나, 자료 수집이 용이하지 않은 경우 재원일수(LOS)를 사용하는 국가가 많다. 심평원의 경우 청구데이터 전수 수집이 가능하므로 진료비(charge)를 이용하여 분석하고 있으며 자료 분석시 재원일수에 관한 사항도 동시에 고려하고 있다. 진료비 청구자료 분석시 비용과 재원일수 모두 중요한 변수이므로 bivariate 방법으로 두 변수 동시에 열외군 제거를 적용하는 방안도 고려할 수 있다(건강보험심사평가원, 2013).

ADRG별 질병군 기준 진료비 분석 결과, 열외비율 5%를 초과하는 질병군이 전체 질병군의 49%를 차지하며, 10%를 초과한 경우도 저빈도 질병군들을 포함하여 3.6%에 달한다. 질병군들의 열외군 비율을 고려하기 위해 현행 단계적 열외방법 적용에서 1단계(식3) 또는 2단계 중 1개의 방법을 선택 적용하는 방안 검토가 필요하다(건강보험심사평가원, 2015).

보건의료재정청(Health Care Financing Administration, HCFA)은 outlier로 알려진 비정상적으로 높은 비용의 사례를 확인하기 위해 산술평균이 아닌 기하평균에서 표준편차를 계산하는 방식(식4)을 사용하였다(Daniel R, 1989).

ADRG별 적정 수준 열외군 비율을 유지하기 위해서는 분포에 맞는 이상치

제거 방법 적용이 타당하며, 이상치 제거 방법 결정에 있어 이상치 제거 목적 (noise, cluster 개념)에 따라 각기 다른 방안을 모색해 볼 수 있으나, 현행방법 이외 실행 가능한 이상치 제거 방법 중 하나로 MAD(Median Absolute Deviation) 방법(식5)과 자연로그변환 방법(식6)을 들 수 있다(건강보험심사평가원, 2015).

$$(식3) \quad Q1 - 3 \times (Q3 - Q1), \quad Q3 + 3 \times (Q3 - Q1)$$

$$(식4) \quad (\text{geo_mean} - 3 \times \text{geo_mean_sd}), (\text{geo_mean} + 3 \times \text{geo_mean_sd})$$

$$(식5) \quad (\text{Median} - 3 \times \text{MAD} \times k), (\text{Median} + 3 \times \text{MAD} \times k), k=1.4826$$

$$(식6) \quad \exp(\ln_mean - 3.0 \times \ln_sd), \exp(\ln_mean + 3.0 \times \ln_sd)$$

기술통계량은 크게 자료가 가지는 분포에 대한 중앙집중성을 알아보는 대푯값과 자료의 흩어진 정도를 측정하는 산포도로 크게 나누어 생각할 수 있다. 중앙집중성을 나타내는 척도로는 평균, 중위수, 그리고 최빈수등을 생각할 수 있다. 또한 산포도의 척도로는 범위, 분산, 표준편차등을 생각할 수 있다.

중앙집중성을 나타내는 값으로써 가장 많이 쓰이는 것이 산술평균이다. 산술평균은 이산형 변수에는 적용할 수 없고, 특히 모든 관측치들의 값이 산술평균을 계산하는데 직접 영향을 미치므로 한 두 개의 이상한 관측값에 대해서도 상당히 영향을 받게된다. 이러한 이상한 관측값을 이상점(extreme value 또는 outlier)이라 하며, 산술평균은 이상점에 의하여 민감하게 영향을 받는 단점이 있다. 반면에 주어진 자료를 크기순으로 배열한 경우 중간에 위치한 값을 중위수라고 한다. 중위수는 주어진 자료들의 크기보다 순위에만 관계하게 되므로 이러한 이상점의 영향을 적게 받는다는 것을 알 수 있다. 이 밖에도 관찰된 자료의 대푯값으로 사용될 수 있는 척도들로는 백분위수(Percentile), 사분위수(Quartile), 범위의 중심값(Mid-range), 기하평균(Geometric mean), 그리고 조화평균(Harmonic mean)등이 있다. 기하평균은 자료가 모두 양수인 경우

에 사용할 수 있다. 일반적으로 기하평균과 조화평균은 변화율과 같은 속성을 가지는 변수의 중앙집중성 척도로 사용된다.

관찰된 자료의 성질을 알아보는 척도로 대푯값 외에 산포도를 생각할 수 있다. 산포도란 관찰된 자료가 대푯값을 중심으로 얼마나 넓게 퍼져 있는가를 알아보는 척도로, 대푯값이 자료의 양적 측면이라면 산포도는 자료의 질적 측면이라고 할 수 있다. 이상점의 영향을 덜 받는 산포도의 척도로 사분위수 범위를 생각할 수 있으며, 산포도로서 가장 많이 사용되는 값으로는 표준편차와 분산이 있다. 한편 분포의 형태를 알아보기 위한 척도로 왜도와 첨도가 있으며, 왜도(skewness)는 분포의 대칭성을 나타내는 척도이고, 첨도(kurtosis)는 분포의 뾰족한 상태를 나타내는 척도이다.

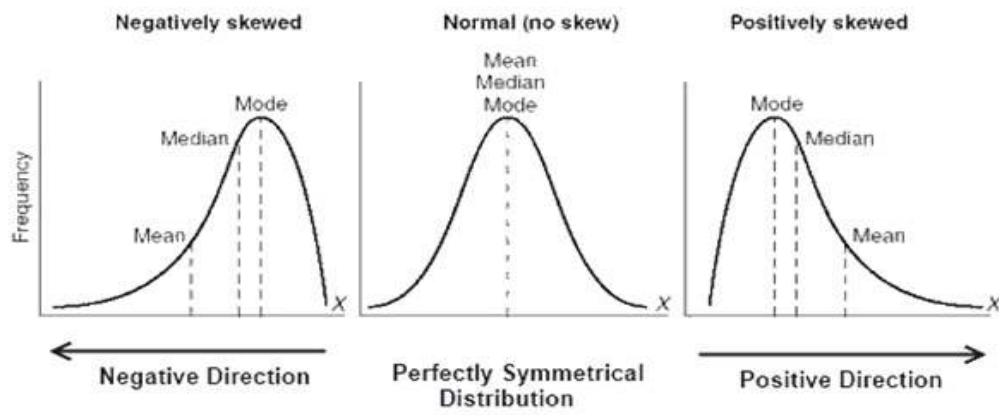


그림 4. 분포에 따른 대표값들의 특징.

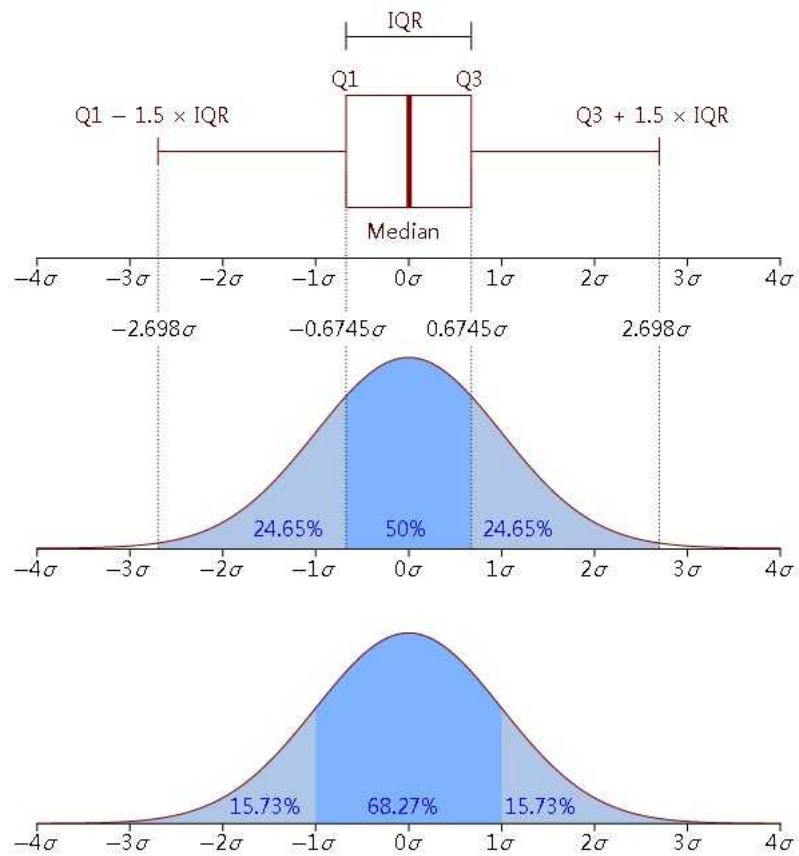


그림 5. $N(0, \sigma^2)$ 인 경우 상자그림과 확률밀도함수.

다. 극단치 처리 방법

2014년 건강보험 행위별 입원청구 전수 자료에서 빈도 수준별 분포의 치우친 정도를 기준으로 총 9개의 질병군을 대상으로 적용하였다. 자료특성에 따른 극단치 처리방법 탐색을 바탕으로 질병군별 다음의 5가지 극단치 처리 방법들을 적용하여 진료비와 재원일수의 trim point가 계산되었다.

(method 1)

$$Q1 - 3 \times (Q3 - Q1), Q3 + 3 \times (Q3 - Q1)$$

(기하평균 - 3 × 표준편차), (기하평균 + 3 × 표준편차)

(method 2)

$$Q1 - 3 \times (Q3 - Q1), Q3 + 3 \times (Q3 - Q1)$$

(method 3)

(기하평균 - 3 × 기하평균 표준편차), (기하평균 + 3 × 기하평균 표준편차)

(method 4)

(Median - 3 × MAD×k), (Median + 3 × MAD×k), k=1.4826

(method 5)

$\exp(\text{로그평균} - 3.0 \times \text{로그표준편차})$, $\exp(\text{로그평균} + 3.0 \times \text{로그표준편차})$

라. 극단치 처리 방법의 평가

질병군 별 진료비와 재원일수에 극단치 처리 방법들을 적용하여 열외군과 비열외군에 대해 상대적으로 가장 큰 일치를 나타내는 방법이 가장 만족스러운 극단치 처리 방법이라는 가설을 두고 있다(Cots F 등, 2003). 질병군별 진료비와 재원일수에 극단치 처리방법들을 적용하여 극단치 판단 결과의 도수를 각 칸에 정리하여 분할표(표5)를 구성하였다. 그리고 극단치의 제거 비율과 accuracy, Kappa coefficient, sensitivity, specificity, ROC curve 지수를 평가 지표로 설정하였다.

표 5. 진료비와 재원일수의 극단치 유무 분할표

LOS (Length of stay)	Charge		
	Outliers	Inliers	Total
Outliers	a	b	a+b
Inliers	c	d	c+d
Total	a+c	b+d	N=a+b+c+d

1) Accuracy

정확도(accuracy) 지수는 전체 환자에서 진료비와 재원일수의 outliers와 inliers가 일치하는 사례의 비율이며, 0과 1사이 값을 가진다.

$$\text{accuracy} = \frac{a+d}{N}$$

2) Kappa coefficient

카파계수(Kappa coefficient)는 우연히 일치할 확률을 고려한 일치도 측정을 나타내는 지수로서 음수와 1사이 값을 가진다. P_o 는 관측된 일치의 사례에 해당하는 비율이며, P_c 는 데이터로부터 계산된 우연히 일치할 사례의 비율이다.

$$\text{Kappa coefficient} = \frac{P_o - P_c}{1 - P_c}$$

$$P_o = \frac{a+d}{N}$$

$$P_c = \frac{((a+b) \times (a+c)) + ((c+d) \times (b+d))}{N^2}$$

3) Sensitivity

민감도(sensitivity) 지수는 환자의 진료비가 outliers인 경우 재원일수에서도 outliers로 나타나는 사례의 비율이며, 0과 1사이의 값을 가진다.

$$\text{sensitivity} = \frac{a}{a+c}$$

4) Specificity

특이도(specificity) 지수는 환자의 진료비가 inliers인 경우 재원일수에서도 inliers로 나타나는 사례의 비율이며, 0과 1사이의 값을 가진다.

$$\text{specificity} = \frac{d}{b+d}$$

5) ROC curve(Receiver operating characteristic curve)

상대작용특성곡선(ROC curve)는 극단치 방법 평가시 절단점에 대하여 민감도와 특이도를 계산한 다음 Y축에는 sensitivity, X축에는 1-specificity를 그래프로 그리는 방법으로 ROC curve 지수는 ROC curve의 아래 면적값(area under the curve:AUC)을 말하며, ROC curve 곡선 아래의 면적은 0과 1사이 값을 가지며 값이 클수록 일치도가 커짐을 보여준다.

$$\text{ROC} = \frac{\text{sensitivity} + \text{specificity}}{2}$$

IV. 연구결과

1. 분석자료의 일반적 특성

KDRG의 성과 측정을 위해 사용한 자료는 건강보험심사평가원에 2014년 1월 1일부터 12월 31일 사이 Electronic Data Interchange(EDI)로 청구된 건강보험 행위별 청구자료 중 입원진료 전체 7,691,707건이다. 그 중 추가 청구건의 경우 누락된 진료내역만 추가로 청구하는 건으로 원 청구와 합산되지 않을 경우 의미가 없으므로 제외하였으며, 입원료 최저금액(보건의료원/낮병동) 미만인 건을 제외하였다. 그리고 주진단 범주 분류에서 진료비 변이가 큰 시술로 이루어진 PreMDC와 Error DRG는 범주 유형 특성상 분석에서 제외하였으며, 요양기관 종별로는 의원, 병원, 종합병원, 상급종합병원을 분석대상에 포함하였다. 이와 같은 데이터의 정제과정을 거쳐 7,407,249건의 자료가 구축되었다.

가. 분포 특성에 따른 극단치 처리 방법별 이상점 시뮬레이션

실제 청구자료 분석에 앞서 분포 특성에 따라 방법별로 이상점을 확인하기 위해 시뮬레이션을 실시하였다. 모집단의 분포가 $N(100000, 1000^2)$ 인 경우와 $LN(10, 1)$ 인 경우 각각 100,000건의 난수를 생성하여 극단치 방법들을 적용한 후 방법별로 이상점을 확인하였다. 모집단의 분포가 완전한 정규분포일 경우 이론적으로 산술평균과 중앙값, 기하평균이 동일한 값을 가지게 되며, 오른쪽으로 꼬리가 긴 형태의 분포에서는 산술평균값이 극단값의 영향을 많이 받게 되므로 중앙값과 기하평균보다 상대적으로 높은 값을 가지게 된다.

시뮬레이션 결과에서 100,000건의 난수를 생성한 결과 정규분포일 경우 극단치 제거 방법 5가지 상·하단의 trim point가 비슷한 값으로 나타났다. 반면, 로그정규분포의 경우에는 상·하단의 trim point의 값이 다양하게 나타났다. 특히 중앙값절대편차를 이용한 method4 적용시 기타 방법들에 비해 상대적으로 trim point가 낮게 나타났으며, 로그변환을 이용하는 method5 적용시 trim point가 가장 높은 값으로 나타났다. 현행 적용중인 method1의 경우 2단계 상·하단 trim point 구간이 1단계 상·하단 trim point 구간에 포함되었으며, 1단계 trim point 값은 상대적으로 높지 않았으나 2단계에서는 trim point 값이 상대적으로 낮게 나타났다.

표 6. 분포 특성에 따른 극단치 처리 방법별 이상점

methods	step	$N(100000, 1000^2)$		$LN(10, 1)$	
		low	high	low	high
method1	step1	95244.39	104750.41	-85399.52	139861.98
	step2	96987.33	102996.69	-60360.32	101103.15
method2	step1	95244.39	104750.41	-85399.52	139861.98
method3	step1	96987.29	102996.73	-124264.88	168186.90
method4	step1	96973.99	103013.70	-36609.31	80392.03
method5	step1	97031.80	103042.52	1088.27	443168.30

표 7. 분포 특성에 따른 극단치 처리 방법별 열외군 비율

methods	step	$N(100000, 1000^2)$			$LN(10, 1)$		
		Inlier	outlier low	outlier high	Inlier	outlier low	outlier high
method1	step1	99.75%	0.13%	0.12%	93.57%	-	6.43%
	step2						
method2	step1	100%	-	-	96.74%	-	3.26%
method3	step1	99.75%	0.13%	0.12%	97.94%	-	2.06%
method4	step1	99.76%	0.12%	0.12%	90.23%	-	9.77%
method5	step1	99.74%	0.15%	0.11%	99.75%	0.12%	0.13%

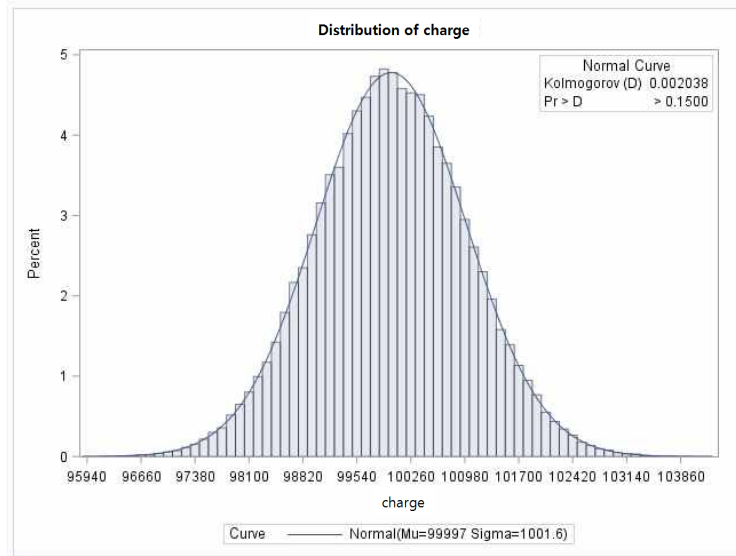


그림 6. 모집단의 분포가 $N(100000, 1000^2)$ 인 경우 히스토그램.

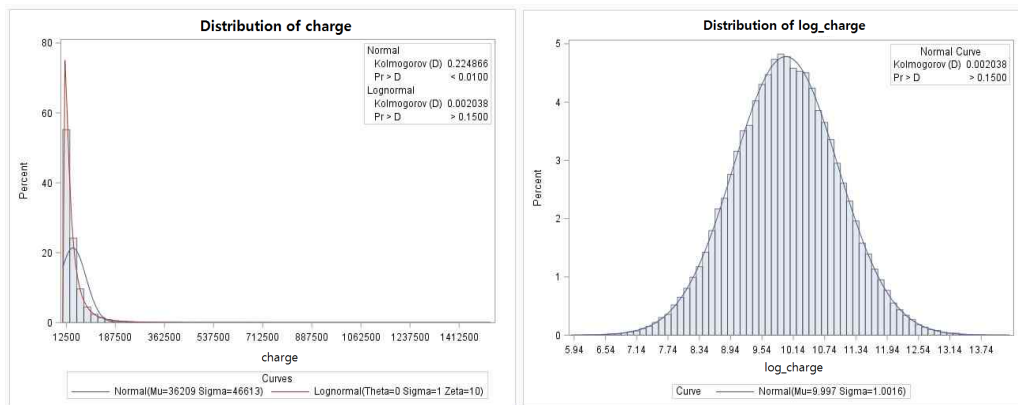


그림 7. 모집단의 분포가 $LN(10, 1)$ 인 경우 히스토그램.

나. MDC 및 ADRG 현황

분석에 이용된 2014년 KDRG Version 4.0의 자료에서 Pre MDC와 Error DRG를 제외한 MDC는 25개이고, ADRG 단계의 그룹수는 1,003개였다. ADRG 수준에서 가장 많은 질병군을 가진 MDC는 'MDC 08 근골격계 및 결합조직의 질환 및 장애'로, 107개의 질병군을 가지고 있어 전체 질병군 개수 ADRG 1,003개 중 10.67%를 차지하였다. 반면에 가장 적은 수의 질병군을 가진 MDC는 'MDC 20 알콜/약물 남용 및 알콜/약물로 인한 기질적 정신 장애'로 총 1개의 질병군으로 이루어져 전체 질병군 개수의 0.10%를 차지하고 있다.

표 8. MDC별 ADRG의 빈도

(unit: N(%))

	MDC	ADRG
01	(B) Diseases and Disorders of the Nervous System	139 (13.86)
02	(C) Diseases and Disorders of the Eye	45 (4.49)
03	(D) Diseases and Disorders of the Ear, Nose, Mouth and Throat	77 (7.68)
04	(E) Diseases and Disorders of the Respiratory System	55 (5.48)
05	(F) Diseases and Disorders of the Circulatory System	99 (9.87)
06	(G) Diseases and Disorders of the Digestive System	83 (8.28)
07	(H) Diseases and Disorders of the Hepatobiliary System and Pancreas	56 (5.58)
08	(I) Diseases and Disorders of the Musculoskeletal System and Connective Tissue	107 (10.67)
09	(J) Diseases and Disorders of the Skin, Subcutaneous Tissue and Breast	34 (3.39)
10	(K) Endocrine, Nutritional and Metabolic Diseases and Disorders	38 (3.79)
11	(L) Diseases and Disorders of the Kidney and Urinary Tract	62 (6.18)
12	(M) Diseases and Disorders of the Male Reproductive System	29 (2.89)
13	(N) Diseases and Disorders of the Female Reproductive System	37 (3.69)
14	(O) Pregnancy, Childbirth and Puerperium	22 (2.19)
15	(P) Newborns and Other Neonates	28 (2.79)

	MDC	ADRG
16 (Q)	Diseases and Disorders of the Blood and Blood-Forming Organs and Immunological Disorders	7 (0.70)
17 (R)	Neoplastic Disorders (Haematological and Solid Neoplasms)	14 (1.40)
18-1 (S)	Infectious and Parasitic Diseases : HIV	5 (0.50)
18-2 (T)	Infectious and Parasitic Diseases	18 (1.79)
19 (U)	Mental Diseases and Disorders	13 (1.30)
20 (V)	Alcohol/Drug Use and Alcohol/Drug Induced Organic Mental Disorders	1 (0.10)
21-1 (W)	Multiple Trauma	6 (0.60)
21-2 (X)	Injuries, Poisoning and Toxic Effects of Drugs	16 (1.60)
22 (Y)	Burns	5 (0.50)
23 (Z)	Factors Influencing Health Status and Other Contacts with Health Services	7 (0.70)
Total		1,003 (100.00)

Note) Pre MDC & Error DRGs(9600,9610,9620,9630,9900,9990) were excluded.

다. 분석대상 질병군

자료의 특성상 질병군별 빈도와 분포가 다양하게 나타나므로 질병군들의 대표적인 특성과 현 상황을 반영하기 위해 빈도와 분포를 고려하여 9개의 질병군을 추출하였다. 빈도를 기준으로 25percentile이하, 25percentile 초과 75percentile 미만, 75percentile 이상 세종류의 그룹으로 나누었다. 그리고 세 그룹별로 진료비의 왜도값이 25 percentile, median, 75percentile에 위치하며, 모든 극단치 방법들을 적용하였을 때 진료비와 재원일수에서 극단치가 발생하는 질병군을 대상으로 하여 9개의 질병군이 선정되었다.

표 9. 분석대상 9개 질병군 ADRG

group_n	group_skewness	ADRG	ADRG Description
25percentile 이하	25percentile	F091	심부정맥혈전증 경피적 수술
	median	B050	뇌기저부 수술
	75percentile	N022	개복에 의한 주요 자궁절제술(악성종양 제외)
25percentile 초과 75percentile 미만	25percentile	R020	기타 신생물 질환(주요 수술 시행)
	median	E693	기타 흉막질환
	75percentile	F704	기타 심근병증
75percentile 이상	25percentile	O620	절박유산
	median	F770	실신 및 허탈
	75percentile	L643	하부 요로 감염

질병군의 빈도와 분포를 고려하여 추출된 9개의 질병군의 빈도는 102건부터 31,609건의 범위를 나타냈으며, 왜도 값은 양의 값을 가지는 1.595부터 11.867로 진료비 자료의 특성에 맞게 우측으로 꼬리가 긴 분포를 나타내는 값을 보였다.

표 10. 9개 질병군의 빈도 및 왜도

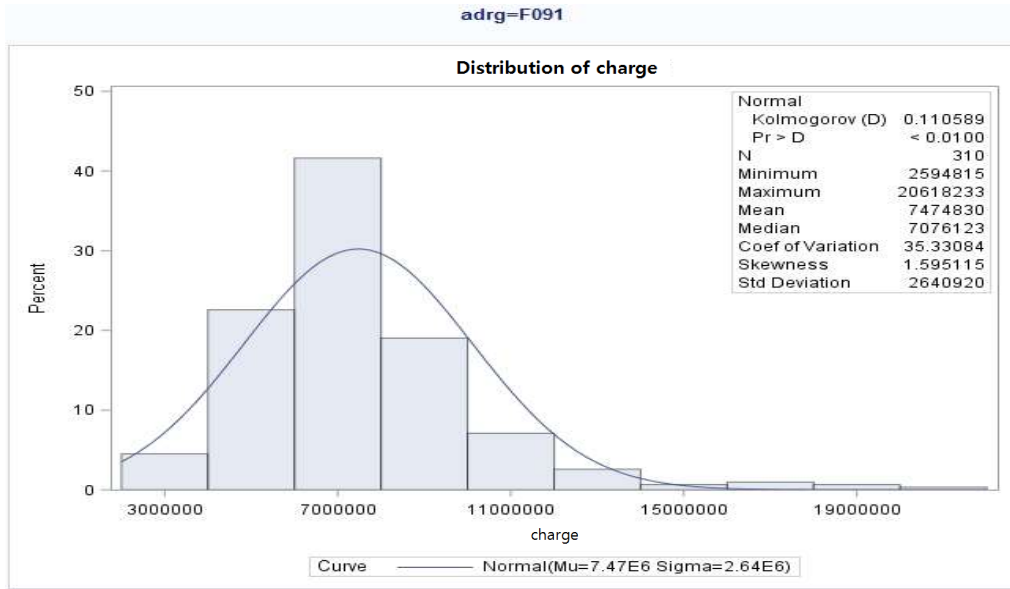
ADRG	ADRG Description	n	skewness
F091	심부정맥혈전증 경피적 수술	310	1.595
B050	뇌기저부 수술	256	2.535
N022	개복에 의한 주요 자궁절제술(악성종양 제외)	102	3.674
R020	기타 신생물 질환(주요 수술 시행)	2,432	3.399
E693	기타 흉막질환	350	5.006
F704	기타 심근병증	427	7.736
O620	절박유산	10,419	5.067
F770	실신 및 허탈	9,245	7.941
L643	하부 요로 감염	31,609	11.867

2. 분석대상 질병군의 진료비 분포 및 재원일수 관계 분석

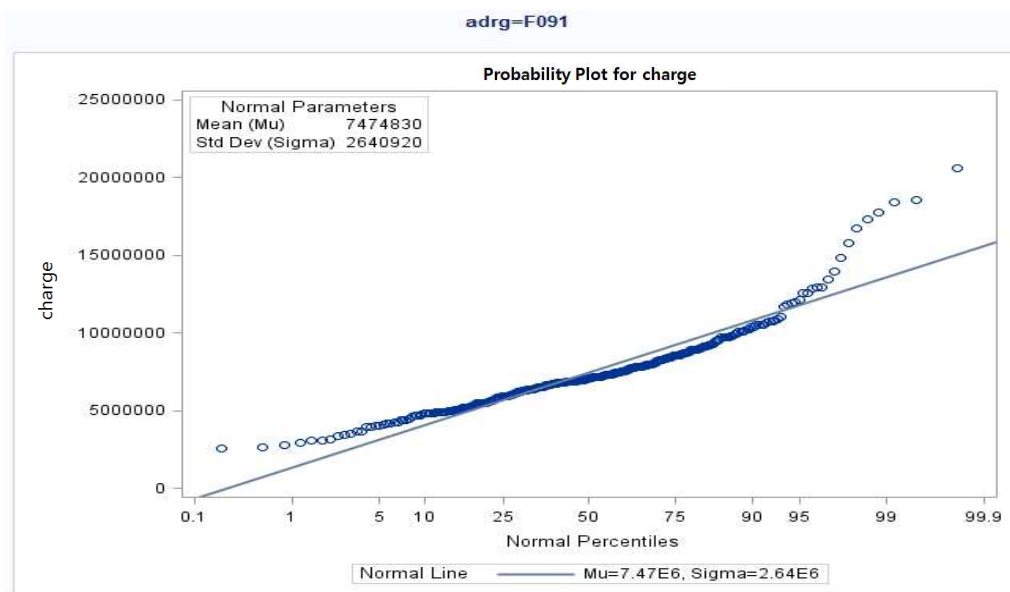
가. 질병군의 진료비 분포

건강보험 청구자료의 진료비 분포를 확인하기 위하여 각 9개 질병군별 진료비의 분포를 히스토그램으로 나타내었으며 정규분포를 가정한 곡선을 함께 나타내었다. 진료비 분포를 확인한 결과, 각 질병군의 분포는 전체적으로 오른쪽으로 꼬리가 긴 양의 왜도(positively skewed)를 보였다.

질병군 진료비 분포의 정규 적합도를 확인하기 위해 Kolmogorove-Smirnov test를 시행한 결과 9개 질병군의 진료비 분포는 모두 정규분포를 따르지 않았다($p < 0.05$). 히스토그램의 형태를 보아도 정규분포의 형태가 아닌 치우친 형태를 보였고, probability plot에서도 진료비를 나타내는 점들이 직선으로부터 이탈되었으며 극단적인 값들이 존재함을 확인할 수 있었다. 이와 같은 내용을 아래의 그림 8~16에 제시하였다.

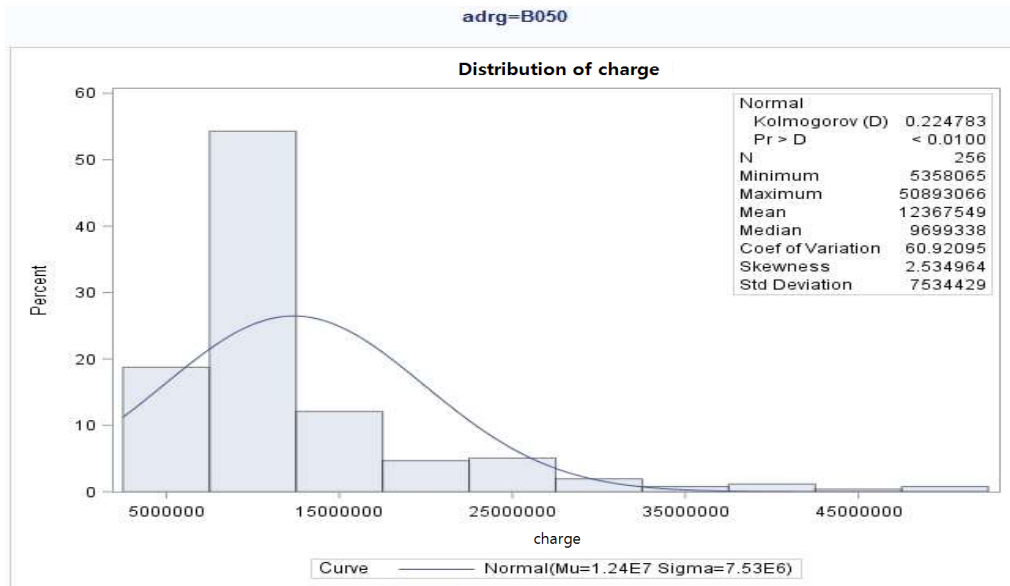


[Histogram]

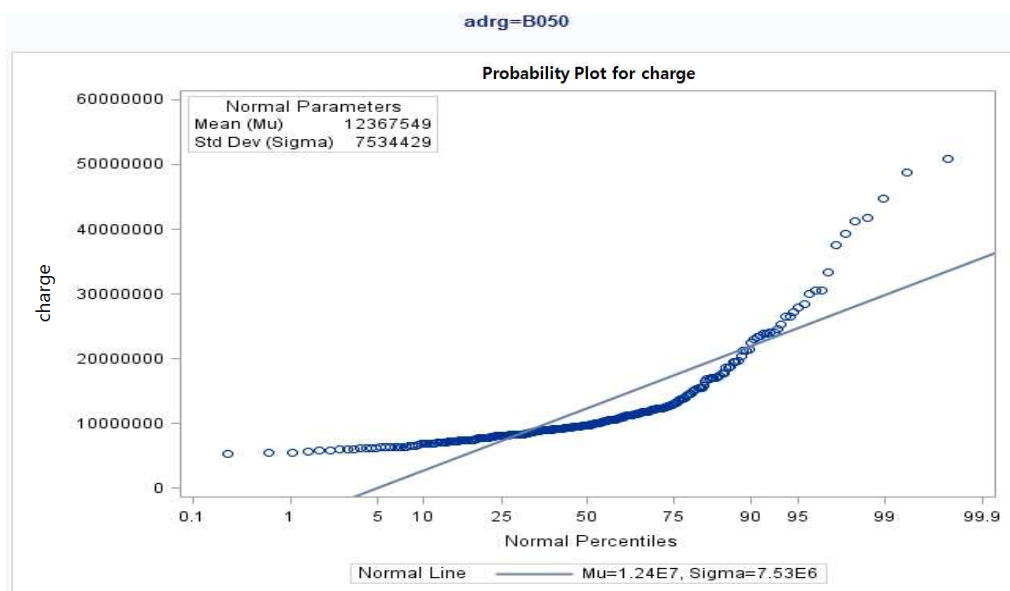


[probability plot]

그림 8. 질병군 F091 histogram & probability plot.

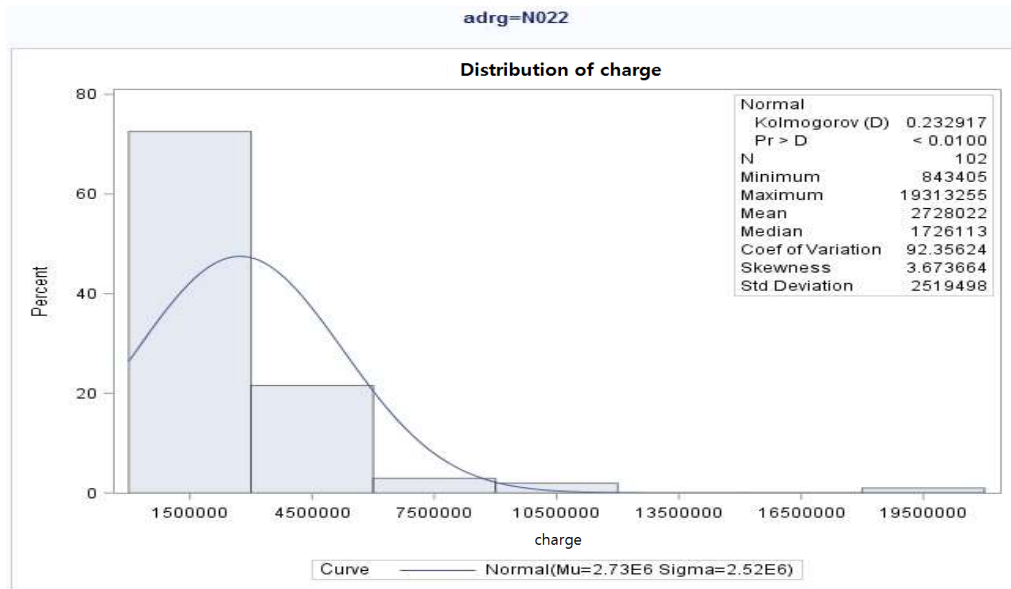


[Histogram]

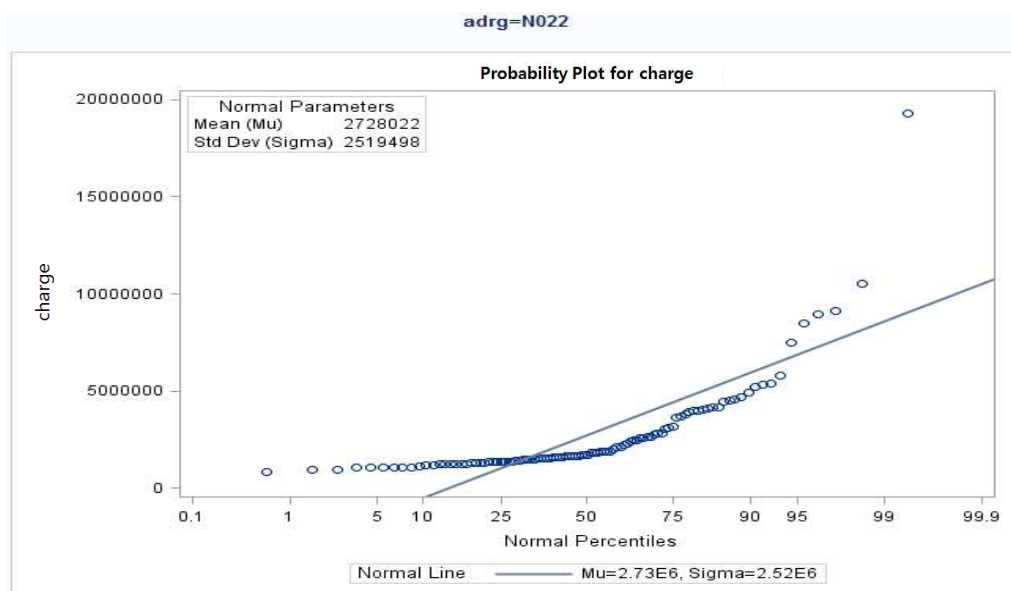


[probability plot]

그림 9. 질병군 B050 histogram & probability plot.

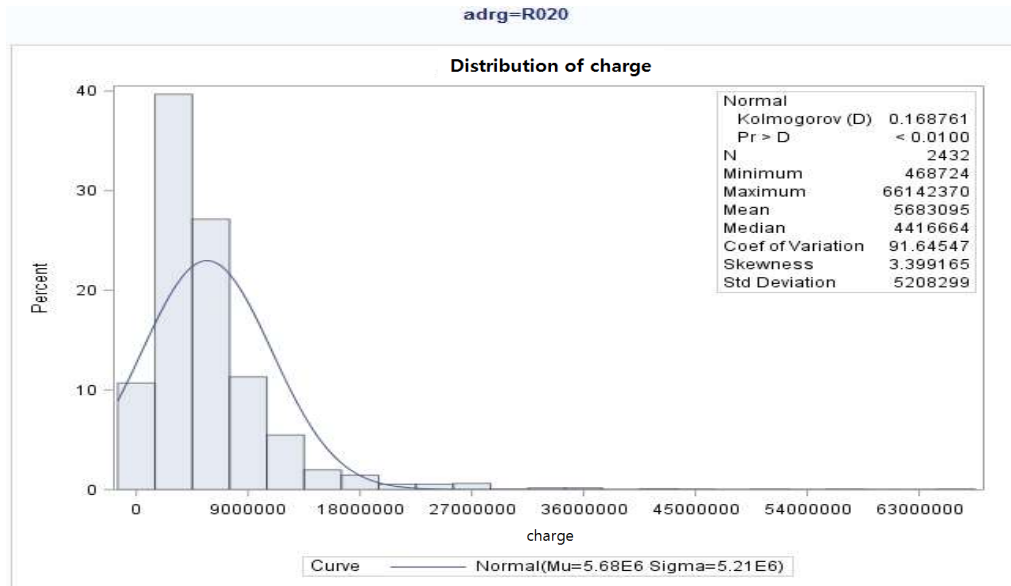


[Histogram]

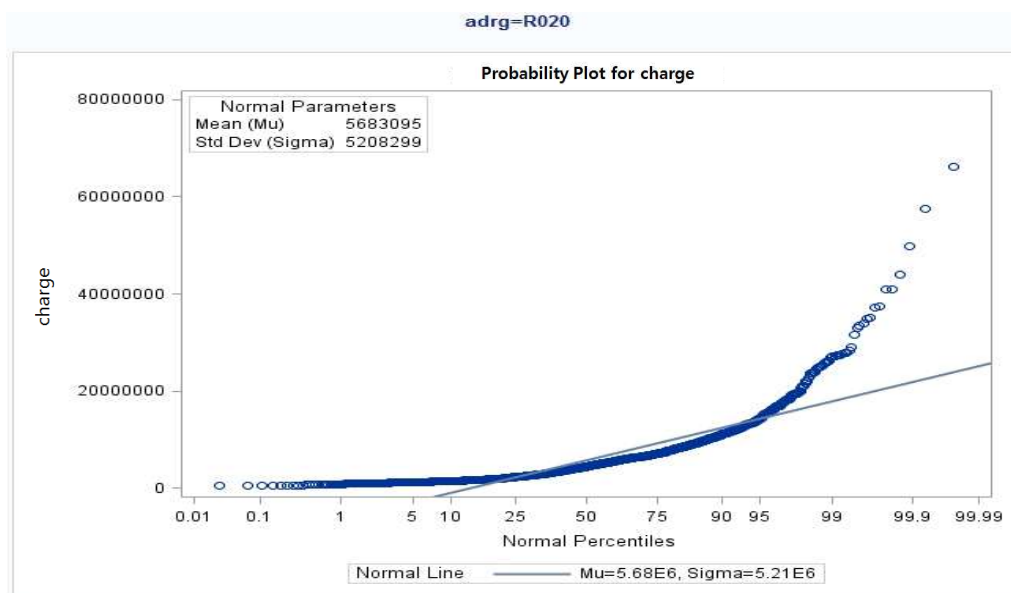


[probability plot]

그림 10. 질병군 N022 histogram & probability plot.

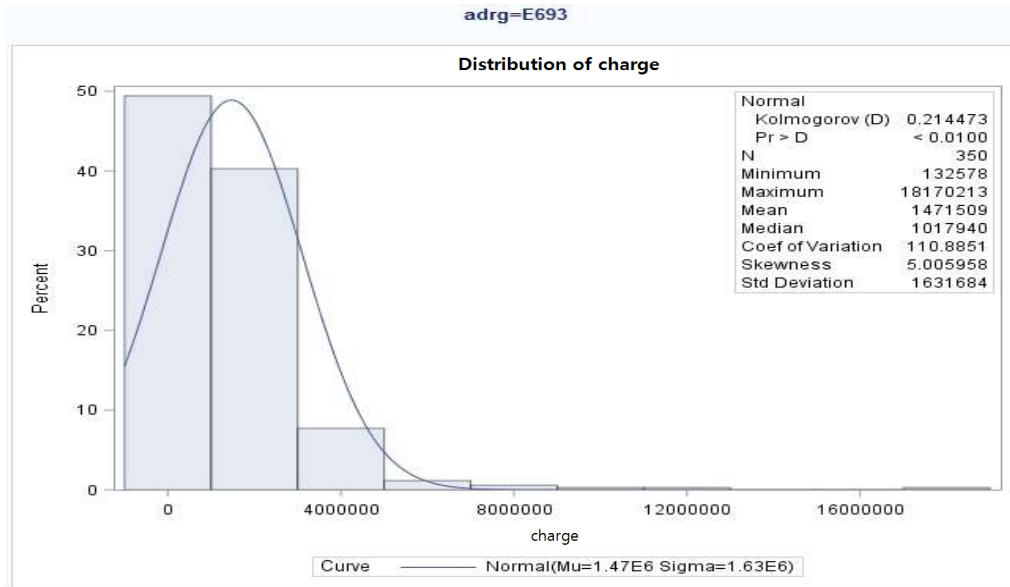


[Histogram]

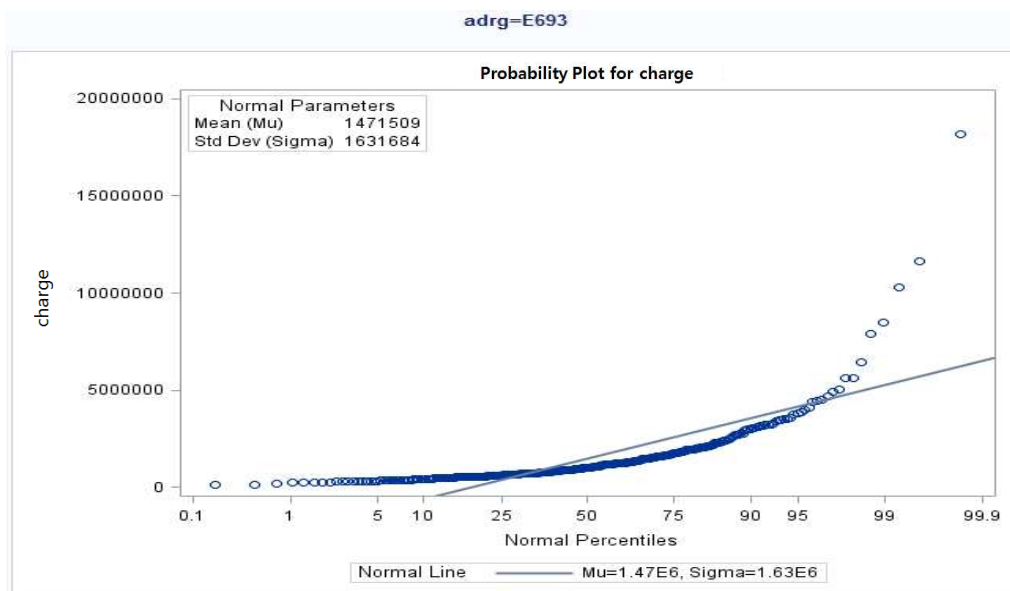


[probability plot]

그림 11. 질병군 R020 histogram & probability plot.

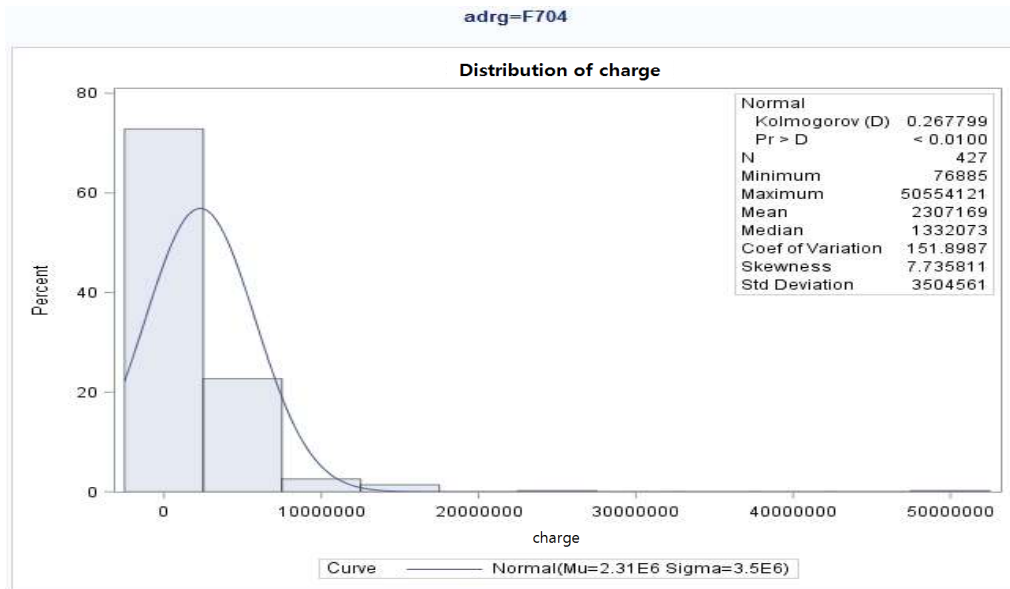


[Histogram]

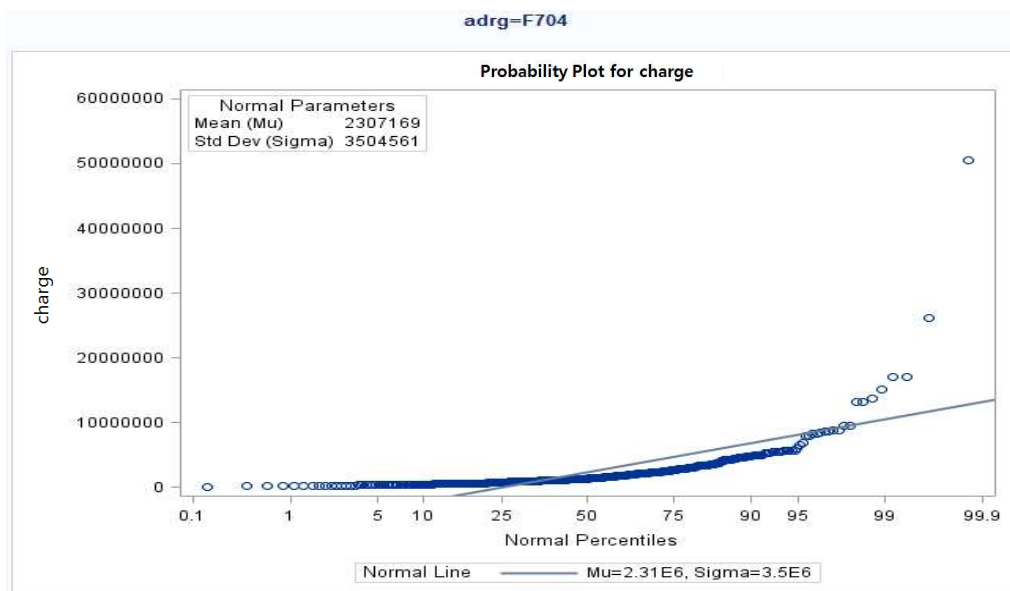


[probability plot]

그림 12. 질병군 E693 histogram & probability plot.

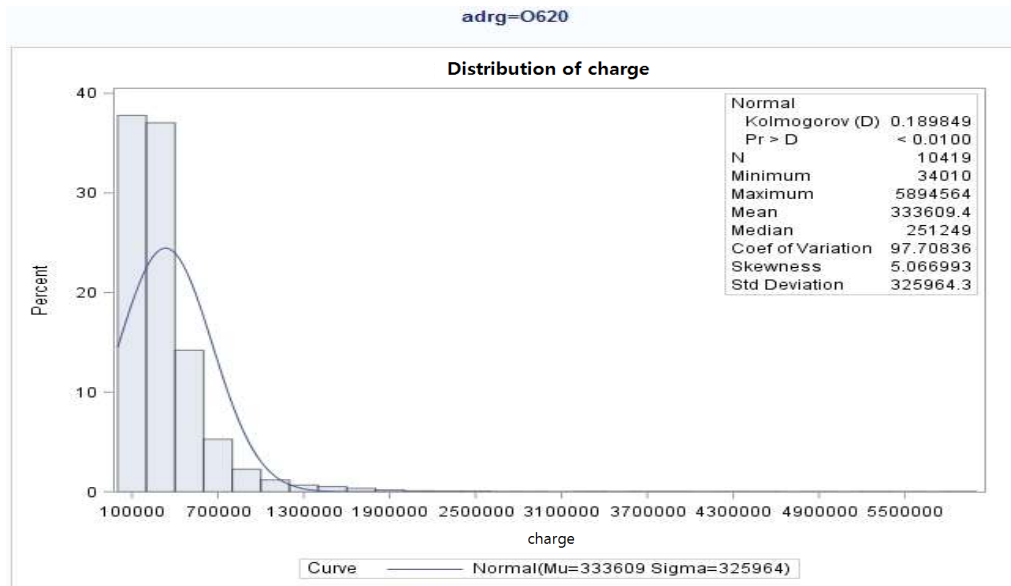


[Histogram]

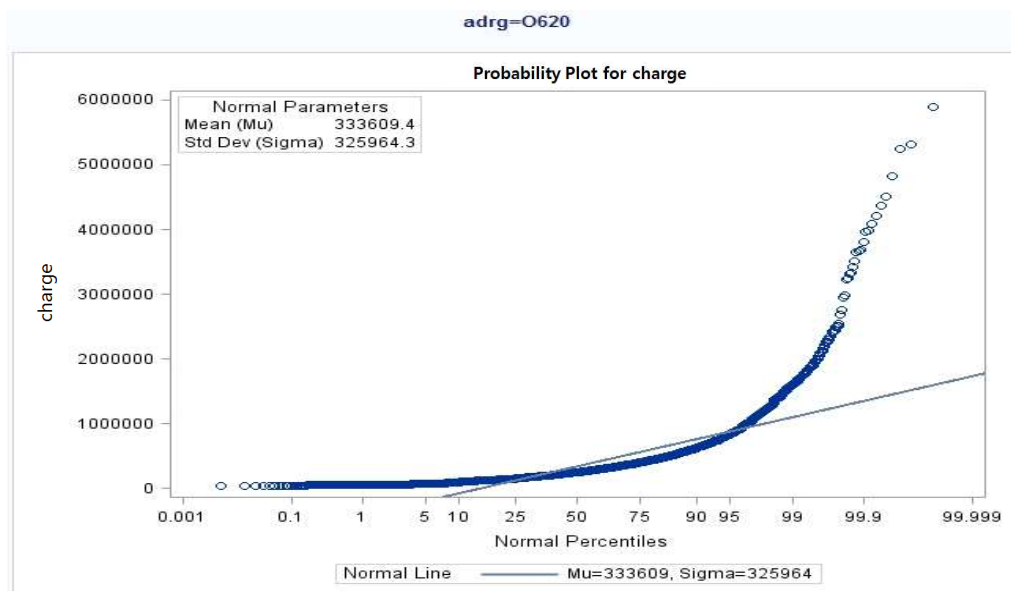


[probability plot]

그림 13. 질병군 F704 histogram & probability plot.

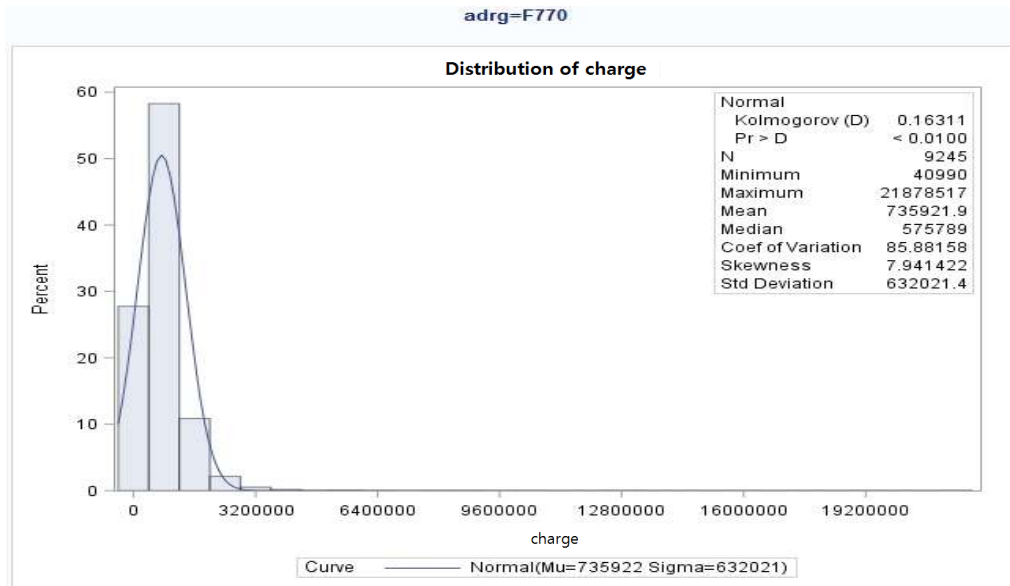


[Histogram]

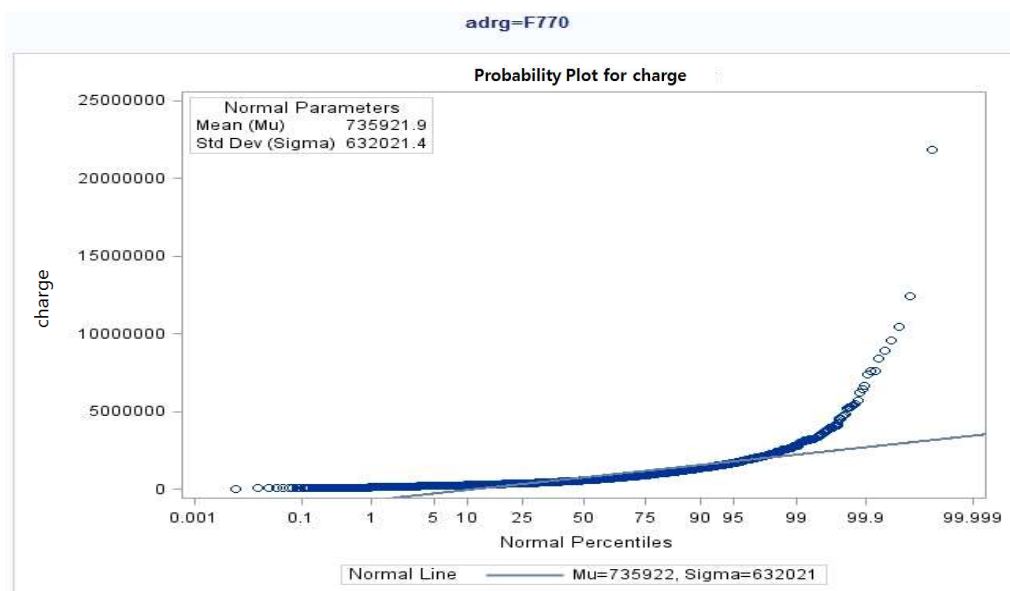


[probability plot]

그림 14. 질병군 O620 histogram & probability plot.

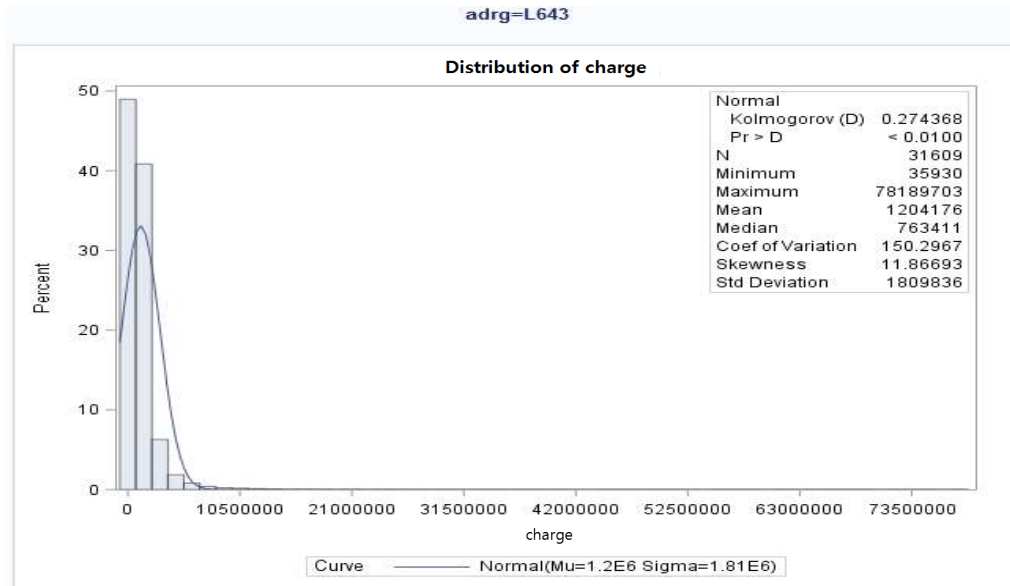


[Histogram]

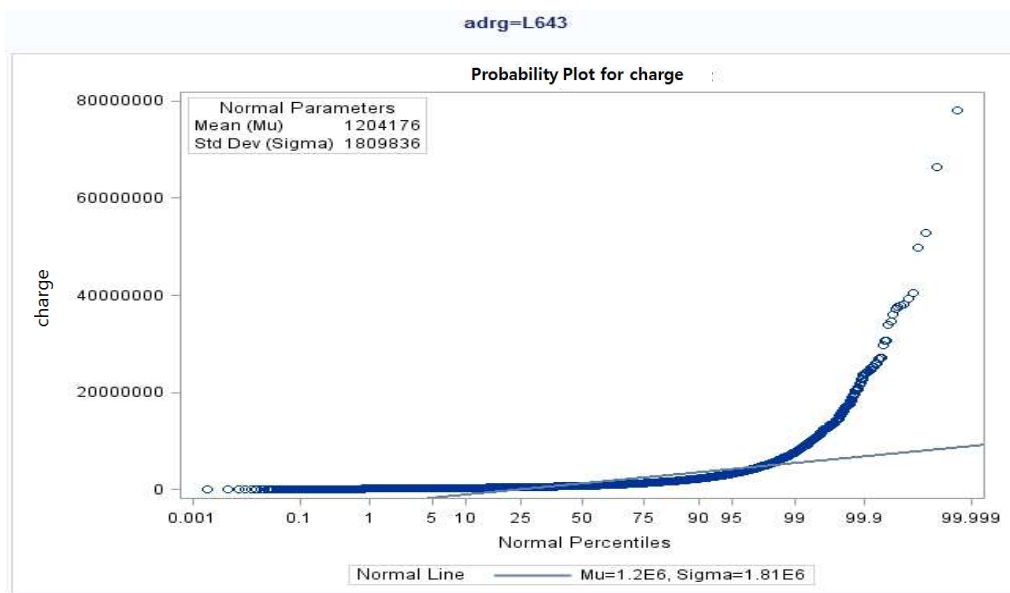


[probability plot]

그림 15. 질병군 F770 histogram & probability plot.



[Histogram]



[probability plot]

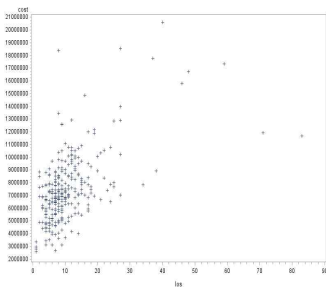
그림 16. 질병군 L643 histogram & probability plot.

나. 질병군의 진료비와 재원일수의 상관관계

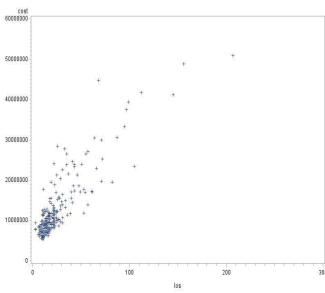
분석대상 질병군별 진료비와 재원일수 두 변수간의 관련성을 선형적인 강도를 통해 알아보기 위해 산점도와 함께 피어슨 상관분석을 실시하였다. 먼저 그림 17의 산점도를 통해 질병군별 진료비와 재원일수가 서로 양의 상관관계가 있음을 파악할 수 있었다. 그리고 피어슨 상관분석을 통해 상관계수와 유의확률 값을 산출한 결과(표11) 9개 질병군의 피어슨상관계수 범위는 0.57~0.93으로 진료비와 재원일수간 양의 상관관계가 있음을 보여준다. 또한 9개 모든 질병군에서 유의확률 값이 유의수준 0.05보다 작은 값으로 나타나 진료비와 재원일수의 관계가 통계적으로 의미가 있는 것으로 나타났다.

표 11. 질병군별 진료비와 재원일수 상관분석

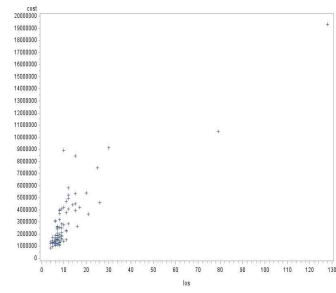
ADRG	Pearson Correlation Coefficients	p-value
F091	0.56779	<0.0001
B050	0.87161	<0.0001
N022	0.85138	<0.0001
R020	0.79229	<0.0001
E693	0.75077	<0.0001
F704	0.76751	<0.0001
O620	0.92610	<0.0001
F770	0.72855	<0.0001
L643	0.82969	<0.0001



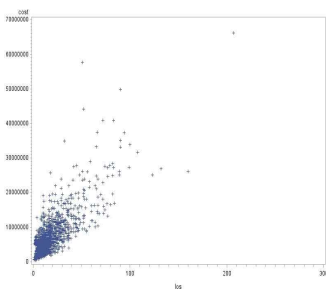
F091



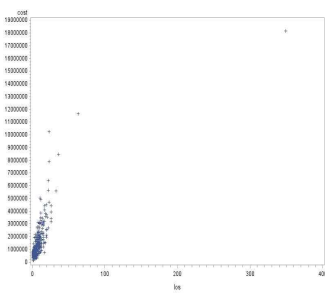
B050



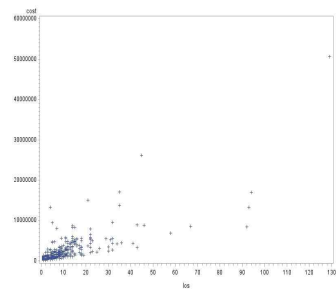
N022



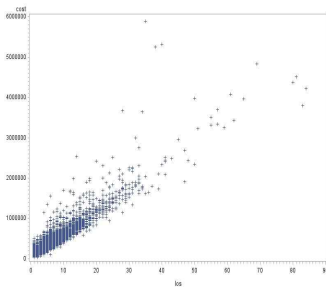
R020



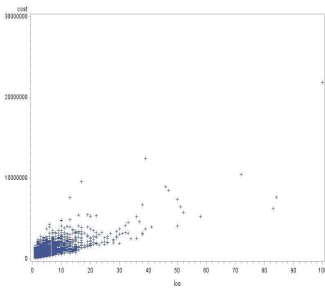
E693



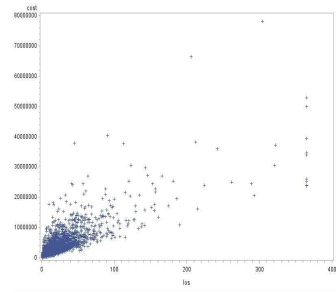
F704



O620



F770



L643

그림 17. 질병군별 진료비와 재원일수 산점도.

3. KDRG 극단치 처리방법의 적정성 분석

KDRG의 극단치 처리방법의 적정성을 분석하기 위해 극단치 처리 방법별로 열외군과 비열외군에 대해 상대적으로 가장 큰 일치(agreement)를 나타내는 방법이 가장 만족스러운 극단치 처리 방법이라는 가설을 설정하였다. 그리고 9개의 질병군 별로 5가지의 극단치 처리 방법(표12)들을 적용한 후 분할표와 평가지표 결과를 산출하여 표 13~21에 제시하였다.

모든 평가 지표에 대해 상대비교한 결과 저빈도군에 속하는 질병군인 F091, B050, N022에서 method3이 가장 상대적으로 높은 일치정도를 보였고, 나머지 6개 질병군에서는 평가 지표들의 높고 낮음이 다양해 질병군마다 일치도가 높은 방법들이 모두 다르게 나타났다.

진료비 기준 극단치로 판단되는 건의 비율이 질병군 전체의 5%가 넘어가는 경우는 method1과 method4를 적용할 경우에서만 발생했다. 재원일수 기준에서도 마찬가지로 method1과 method4를 적용할 경우 대부분의 질병군에서 극단치의 비율이 5%가 초과되었으며, B050 질병군에서 method2를 적용할 시 재원일수 극단건 비율이 5.469%로 5%를 초과하였다. 특히 method4의 경우에는 진료비와 재원일수 기준 모두 10%를 초과하는 경우도 발생하였다.

표 12. KDRG 자료특성에 따른 극단치 처리 방법

method number	method used
method 1	step1. $Q1 - 3 \times (Q3 - Q1)$, $Q3 + 3 \times (Q3 - Q1)$ step2. $\text{geo_mean} - 3 \times \text{sd}$, $\text{geo_mean} + 3 \times \text{sd}$
method 2	$Q1 - 3 \times (Q3 - Q1)$, $Q3 + 3 \times (Q3 - Q1)$
method 3	$(\text{geo_mean} - 3 \times \text{geo_mean_sd})$, $(\text{geo_mean} + 3 \times \text{geo_mean_sd})$
method 4	$(\text{median} - 3 \times \text{MAD} \times k)$, $(\text{median} + 3 \times \text{MAD} \times k)$, $k=1.4826$
method 5	$\exp(\ln_mean - 3 \times \ln_sd)$, $\exp(\ln_mean + 3 \times \ln_sd)$

표 13. 질병군 F091 극단치 방법별 분할표 및 평가지수 비교 결과

질병군 F091 극단치 방법별 2×2 분할표

(method 1)

LOS	charge		
	Outliers	Inliers	Total
Outliers	7	10	17
Inliers	2	291	293
Total	9	301	310

(method 2)

LOS	charge		
	Outliers	Inliers	Total
Outliers	4	5	9
Inliers	2	299	301
Total	6	304	310

(method 3)

LOS	charge		
	Outliers	Inliers	Total
Outliers	5	3	8
Inliers	2	300	302
Total	7	303	310

(method 4)

LOS	charge		
	Outliers	Inliers	Total
Outliers	7	14	21
Inliers	3	286	289
Total	10	300	310

(method 5)

LOS	charge		
	Outliers	Inliers	Total
Outliers	1	5	6
Inliers	1	303	304
Total	2	308	310

질병군 F091 극단치 방법별 평가지수 비교 결과

method	accuracy	Kappa	sensitivity	specificity	ROC curve	charge_outlier	LOS_outlier
method1	0.961	0.520	0.778	0.967	0.872	2.903%	5.484%
method2	0.977	0.522	0.667	0.984	0.825	1.935%	2.903%
method3	0.984	0.658	0.714	0.990	0.852	2.258%	2.581%
method4	0.945	0.427	0.700	0.953	0.827	3.226%	6.774%
method5	0.981	0.243	0.500	0.984	0.742	0.645%	1.935%

표 14. 질병군 B050 극단치 방법별 분할표 및 평가지수 비교 결과

질병군 B050 극단치 방법별 2×2 분할표

(method 1)

LOS	charge		
	Outliers	Inliers	Total
Outliers	14	8	22
Inliers	3	231	234
Total	17	239	256

(method 2)

LOS	charge		
	Outliers	Inliers	Total
Outliers	10	4	14
Inliers	2	240	242
Total	12	244	256

(method 3)

LOS	charge		
	Outliers	Inliers	Total
Outliers	6	2	8
Inliers	1	247	248
Total	7	249	256

(method 4)

LOS	charge		
	Outliers	Inliers	Total
Outliers	19	9	28
Inliers	12	216	228
Total	31	225	256

(method 5)

LOS	charge		
	Outliers	Inliers	Total
Outliers	2	1	3
Inliers	1	252	253
Total	3	253	256

질병군 B050 극단치 방법별 평가지수 비교 결과

method	accuracy	Kappa	sensitivity	specificity	ROC curve	charge_outlier	LOS_outlier
method1	0.957	0.695	0.824	0.967	0.895	6.641%	8.594%
method2	0.977	0.757	0.833	0.984	0.908	4.688%	5.469%
method3	0.988	0.794	0.857	0.992	0.925	2.734%	3.125%
method4	0.918	0.598	0.613	0.960	0.786	12.109%	10.938%
method5	0.992	0.663	0.667	0.996	0.831	1.172%	1.172%

표 15. 질병군 N022 극단치 방법별 분할표 및 평가지수 비교 결과

질병군 N022 극단치 방법별 2×2 분할표

(method 1)

LOS	charge		
	Outliers	Inliers	Total
Outliers	4	3	7
Inliers	2	93	95
Total	6	96	102

(method 2)

LOS	charge		
	Outliers	Inliers	Total
Outliers	3	2	5
Inliers	1	96	97
Total	4	98	102

(method 3)

LOS	charge		
	Outliers	Inliers	Total
Outliers	2	0	2
Inliers	0	100	100
Total	2	100	102

(method 4)

LOS	charge		
	Outliers	Inliers	Total
Outliers	7	1	8
Inliers	11	83	94
Total	18	84	102

(method 5)

LOS	charge		
	Outliers	Inliers	Total
Outliers	1	1	2
Inliers	0	100	100
Total	1	101	102

질병군 N022 극단치 방법별 평가지수 비교 결과

method	accuracy	Kappa	sensitivity	specificity	ROC curve	charge_outlier	LOS_outlier
method1	0.951	0.589	0.667	0.969	0.818	5.882%	6.863%
method2	0.971	0.651	0.750	0.980	0.865	3.922%	4.902%
method3	1.000	1.000	1.000	1.000	1.000	1.961%	1.961%
method4	0.882	0.482	0.389	0.988	0.688	17.647%	7.843%
method5	0.990	0.662	1.000	0.990	0.995	0.980%	1.961%

표 16. 질병군 R020 극단치 방법별 분할표 및 평가지수 비교 결과

질병군 R020 극단치 방법별 2×2 분할표

(method 1)

LOS	charge		
	Outliers	Inliers	Total
Outliers	78	70	148
Inliers	31	2,253	2,284
Total	109	2,323	2,432

(method 2)

LOS	charge		
	Outliers	Inliers	Total
Outliers	38	41	79
Inliers	9	2,344	2,353
Total	47	2,385	2,432

(method 3)

LOS	charge		
	Outliers	Inliers	Total
Outliers	34	24	58
Inliers	17	2,357	2,374
Total	51	2,384	2,432

(method 4)

LOS	charge		
	Outliers	Inliers	Total
Outliers	87	97	184
Inliers	28	2,220	2,248
Total	115	2,317	2,432

(method 5)

LOS	charge		
	Outliers	Inliers	Total
Outliers	1	3	4
Inliers	3	2,425	2,428
Total	4	2,428	2,432

질병군 R020 극단치 방법별 평가지수 비교 결과

method	accuracy	Kappa	sensitivity	specificity	ROC curve	charge_outlier	LOS_outlier
method1	0.958	0.586	0.716	0.970	0.843	4.482%	6.086%
method2	0.979	0.593	0.809	0.983	0.896	1.933%	3.248%
method3	0.983	0.615	0.667	0.990	0.828	2.097%	2.385%
method4	0.949	0.556	0.757	0.958	0.857	4.729%	7.566%
method5	0.998	0.249	0.250	0.999	0.624	0.164%	0.164%

표 17. 질병군 E693 극단치 방법별 분할표 및 평가지수 비교 결과

질병군 E693 극단치 방법별 2×2 분할표

(method 1)

LOS	charge		
	Outliers	Inliers	Total
Outliers	11	3	14
Inliers	6	330	336
Total	17	333	350

(method 2)

LOS	charge		
	Outliers	Inliers	Total
Outliers	4	0	4
Inliers	4	342	346
Total	8	342	350

(method 3)

LOS	charge		
	Outliers	Inliers	Total
Outliers	1	0	1
Inliers	5	344	349
Total	6	344	350

(method 4)

LOS	charge		
	Outliers	Inliers	Total
Outliers	15	4	19
Inliers	14	317	331
Total	29	321	350

(method 5)

LOS	charge		
	Outliers	Inliers	Total
Outliers	1	0	1
Inliers	1	348	349
Total	2	348	350

질병군 E693 극단치 방법별 평가지수 비교 결과

method	accuracy	Kappa	sensitivity	specificity	ROC curve	charge_outlier	LOS_outlier
method1	0.974	0.696	0.647	0.991	0.819	4.857%	4.000%
method2	0.989	0.662	0.500	1.000	0.750	2.286%	1.143%
method3	0.986	0.282	0.167	1.000	0.583	1.714%	0.286%
method4	0.949	0.599	0.517	0.988	0.752	8.286%	5.429%
method5	0.997	0.665	0.500	1.000	0.750	0.571%	0.286%

표 18. 질병군 F704 극단치 방법별 분할표 및 평가지수 비교 결과

질병군 F704 극단치 방법별 2×2 분할표

(method 1)

LOS	charge		
	Outliers	Inliers	Total
Outliers	12	12	24
Inliers	10	393	403
Total	22	405	427

(method 2)

LOS	charge		
	Outliers	Inliers	Total
Outliers	8	3	11
Inliers	8	408	416
Total	16	411	427

(method 3)

LOS	charge		
	Outliers	Inliers	Total
Outliers	3	4	7
Inliers	5	415	420
Total	8	419	420

(method 4)

LOS	charge		
	Outliers	Inliers	Total
Outliers	15	10	25
Inliers	29	373	402
Total	44	383	427

(method 5)

LOS	charge		
	Outliers	Inliers	Total
Outliers	1	0	1
Inliers	2	424	426
Total	3	424	427

질병군 F704 극단치 방법별 평가지수 비교 결과

method	accuracy	Kappa	sensitivity	specificity	ROC curve	charge_outlier	LOS_outlier
method1	0.948	0.495	0.545	0.970	0.758	5.152%	5.621%
method2	0.974	0.580	0.500	0.993	0.746	3.747%	2.576%
method3	0.979	0.389	0.375	0.990	0.683	1.874%	1.639%
method4	0.909	0.389	0.341	0.974	0.657	10.304%	5.855%
method5	0.995	0.498	0.333	1.000	0.667	0.703%	0.234%

표 19. 질병군 O620 극단치 방법별 분할표 및 평가지수 비교 결과

질병군 O620 극단치 방법별 2×2 분할표

(method 1)

LOS	charge		
	Outliers	Inliers	Total
Outliers	416	119	535
Inliers	102	9,782	9,884
Total	518	9,901	10,419

(method 2)

LOS	charge		
	Outliers	Inliers	Total
Outliers	192	23	215
Inliers	66	10,138	10,204
Total	258	10,161	10,419

(method 3)

LOS	charge		
	Outliers	Inliers	Total
Outliers	158	57	215
Inliers	40	10,164	10,204
Total	198	10,221	10,419

(method 4)

LOS	charge		
	Outliers	Inliers	Total
Outliers	521	115	636
Inliers	123	9,660	9,783
Total	644	9,775	10,419

(method 5)

LOS	charge		
	Outliers	Inliers	Total
Outliers	32	28	60
Inliers	6	10,353	10,359
Total	38	10,381	10,419

질병군 O620 극단치 방법별 평가지수 비교 결과

method	accuracy	Kappa	sensitivity	specificity	ROC curve	charge_outlier	LOS_outlier
method1	0.979	0.779	0.803	0.988	0.896	4.972%	5.135%
method2	0.991	0.808	0.744	0.998	0.871	2.476%	2.064%
method3	0.991	0.760	0.798	0.994	0.896	1.900%	2.064%
method4	0.977	0.802	0.809	0.988	0.899	6.181%	6.104%
method5	0.997	0.652	0.842	0.997	0.920	0.365%	0.576%

표 20. 질병군 F770 극단치 방법별 분할표 및 평가지수 비교 결과

질병군 F770 극단치 방법별 2×2 분할표

(method 1)

LOS	charge		
	Outliers	Inliers	Total
Outliers	210	297	507
Inliers	151	8,587	8,738
Total	361	8,884	9,245

(method 2)

LOS	charge		
	Outliers	Inliers	Total
Outliers	90	229	319
Inliers	50	8,876	8,926
Total	140	9,105	9,245

(method 3)

LOS	charge		
	Outliers	Inliers	Total
Outliers	69	88	157
Inliers	65	9,023	9,088
Total	134	9,111	9,245

(method 4)

LOS	charge		
	Outliers	Inliers	Total
Outliers	294	392	686
Inliers	255	8,304	8,559
Total	549	8,696	9,245

(method 5)

LOS	charge		
	Outliers	Inliers	Total
Outliers	16	46	62
Inliers	40	9,143	9,183
Total	56	9,189	9,245

질병군 F770 극단치 방법별 평가지수 비교 결과

method	accuracy	Kappa	sensitivity	specificity	ROC curve	charge_outlier	LOS_outlier
method1	0.952	0.459	0.582	0.967	0.774	3.905%	5.484%
method2	0.970	0.379	0.643	0.975	0.809	1.514%	3.451%
method3	0.983	0.466	0.515	0.990	0.753	1.449%	1.698%
method4	0.930	0.439	0.536	0.955	0.745	5.938%	7.420%
method5	0.991	0.267	0.286	0.995	0.640	0.606%	0.671%

표 21. 질병군 L643 극단치 방법별 분할표 및 평가지수 비교 결과

질병군 L643 극단치 방법별 2×2 분할표

(method 1)

LOS	charge		
	Outliers	Inliers	Total
Outliers	1,291	637	1,928
Inliers	838	28,843	29,681
Total	2,129	29,480	31,609

(method 2)

LOS	charge		
	Outliers	Inliers	Total
Outliers	680	337	1,017
Inliers	553	30,039	30,592
Total	1,233	30,376	31,609

(method 3)

LOS	charge		
	Outliers	Inliers	Total
Outliers	261	92	353
Inliers	196	31,060	31,256
Total	457	31,152	31,609

(method 4)

LOS	charge		
	Outliers	Inliers	Total
Outliers	2,218	1,119	3,337
Inliers	979	27,293	28,272
Total	3,197	28,412	31,609

(method 5)

LOS	charge		
	Outliers	Inliers	Total
Outliers	143	81	224
Inliers	144	31,241	31,385
Total	287	31,322	31,609

질병군 L643 극단치 방법별 평가지수 비교 결과

method	accuracy	Kappa	sensitivity	specificity	ROC curve	charge_outlier	LOS_outlier
method1	0.953	0.612	0.606	0.978	0.792	6.735%	6.100%
method2	0.972	0.590	0.552	0.989	0.770	3.901%	3.217%
method3	0.991	0.640	0.571	0.997	0.784	1.446%	1.117%
method4	0.934	0.642	0.694	0.961	0.827	10.114%	10.557%
method5	0.993	0.556	0.498	0.997	0.748	0.908%	0.709%

4. KDRG 극단치 처리방법 대안 탐색 및 결과

앞 장에서 KDRG 극단치 처리 방법의 적정성에 대하여 분석한 결과, 진료비와 재원일수 극단치의 비율 측면에서는 방법의 종류에 따라 상대적 차이를 보였다. 일치도 측면에서는 평가지표의 다양함으로 인해 저빈도 그룹을 제외한 질병군에서는 어느 한 방법이 일관성 있게 일치도가 높게 나타나지는 않았다. 따라서 본 장에서는 KDRG 극단치 처리방법 대안을 탐색하기 위해 KDRG 현행 극단치 처리방법의 가장 큰 문제점인 극단치 비율 문제를 고려하여 극단치 처리 방법에 따른 진료비와 재원일수의 열외군 비율을 탐색하였다. 또한 일치도 측면 평가지표에서 카파계수의 경우 자료가 비대칭 구조일 때 일치도가 높더라도 카파계수는 낮게 산출되는 한계점을 내포하고 있다. 민감도와 특이도 지수를 계산하여 AUC값을 통해 일치도를 파악하는 ROC curve 지수 값을 일치도 측면 평가지표로 설정하여 상대 비교하였다.

가. KDRG 극단치 처리 방법 대안 탐색

진료비 기준 질병군별 극단치 처리 방법에 따라 열외군의 비율을 보면 method1과 method4를 적용할 때 대부분의 질병군에서 열외군의 비율이 5%를 초과하였다. 특히 method4를 적용할 경우 열외군 비율이 10%를 초과하는 경우도 발생하였다. 재원일수 기준에서도 method1을 적용할 경우 9개 질병군 중 8개 질병군에서 열외군 비율이 5%를 초과하였으며, method4를 적용할 경우 모든 질병군에서 열외군 비율이 5%를 초과하였다. 재원일수 기준 질병군 B050에 method2를 적용할 때 열외군 비율이 5.469%로 5%를 초과하였다. 전

체적으로 method1과 method4를 극단치 처리 방법으로 적용할 시 기타 방법에 비해 상대적으로 열외군의 비율이 높게 나타났다.

열외군 비율이 다른 방법에 비해 상대적으로 높은 method1과 method4를 제외한 method2, method3, method5를 대상으로 하여 열외군과 비열외군에 대해 일치정도를 ROC curve 값으로 상대비교한 결과 method5는 질병군 O620에서만 가장 큰 일치도를 보였고, 질병군 E693에서는 일치도가 0.750으로 methods2와 공동으로 가장 큰 일치도를 보였다. 전체적으로는 method2와 method3이 9개 질병군 중 각각 4개의 질병군에서 가장 큰 일치정도를 보였다. method2와 method3이 상대적으로 열외군 비율은 낮고, 일치도 측면에서는 상대적으로 높은 수치를 보였으며, 그 중에서도 method2 방법이 질병군별 일치도 측면에서 변동의 폭이 작아 상대적으로 안정적이게 나타났다.

표 22. 질병군별 진료비 극단치 처리 방법 적용시 열외군 비율

(unit:%)

ADRG	method1	method2	method3	method4	method5
F091	2.903	1.935	2.258	3.226	0.645
B050	6.641	4.688	2.734	12.109	1.172
N022	5.882	3.922	1.961	17.647	0.980
R020	4.482	1.933	2.097	4.729	0.164
E693	4.857	2.286	1.714	8.286	0.571
F704	5.152	3.747	1.874	10.304	0.703
O620	4.972	2.476	1.900	6.181	0.365
F770	3.905	1.514	1.449	5.938	0.606
L643	6.735	3.901	1.446	10.114	0.908

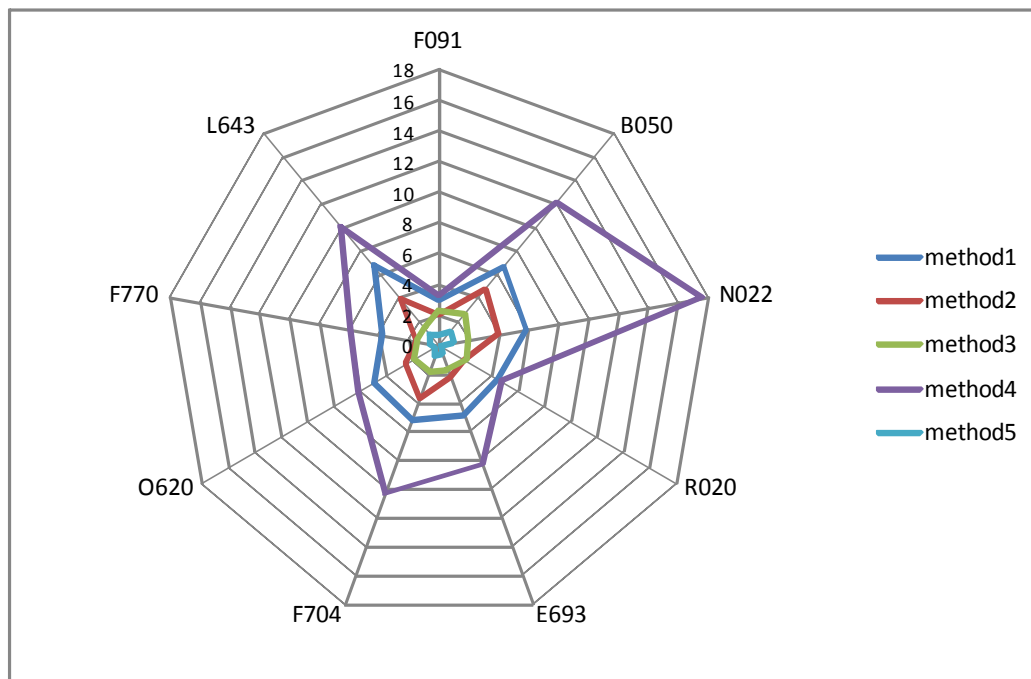


그림 18. 질병군별 진료비 극단치 처리 방법 적용시 열외군 비율.

표 23. 질병군별 재원일수 극단치 처리 방법 적용시 열외군 비율

(unit:%)

ADRG	method1	method2	method3	method4	method5
F091	5.484	2.903	2.581	6.774	1.935
B050	8.594	5.469	3.125	10.938	1.172
N022	6.863	4.902	1.961	7.843	1.961
R020	6.086	3.248	2.385	7.566	0.164
E693	4.000	1.143	0.286	5.429	0.286
F704	5.621	2.576	1.639	5.855	0.234
O620	5.135	2.064	2.064	6.104	0.576
F770	5.484	3.451	1.698	7.420	0.671
L643	6.100	3.217	1.117	10.557	0.709

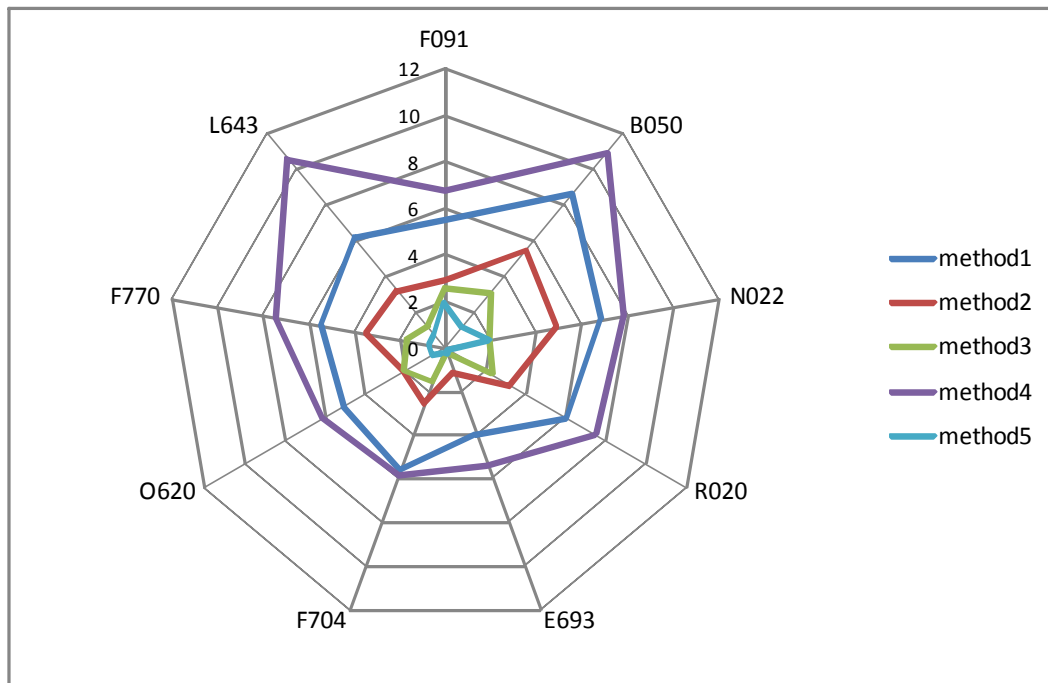


그림 19. 질병군별 재원일수 극단치 처리 방법 적용시 열외군 비율.

표 24. 열외군 비율이 낮은 극단치 처리 방법들 간의 일치도(ROC)

ADRG	method2	method3	method5
F091	0.825	0.852	0.742
B050	0.908	0.925	0.831
N022	0.865	1.000	0.995
R020	0.896	0.828	0.624
E693	0.750	0.583	0.750
F704	0.746	0.683	0.667
O620	0.871	0.896	0.920
F770	0.809	0.753	0.640
L643	0.770	0.784	0.748

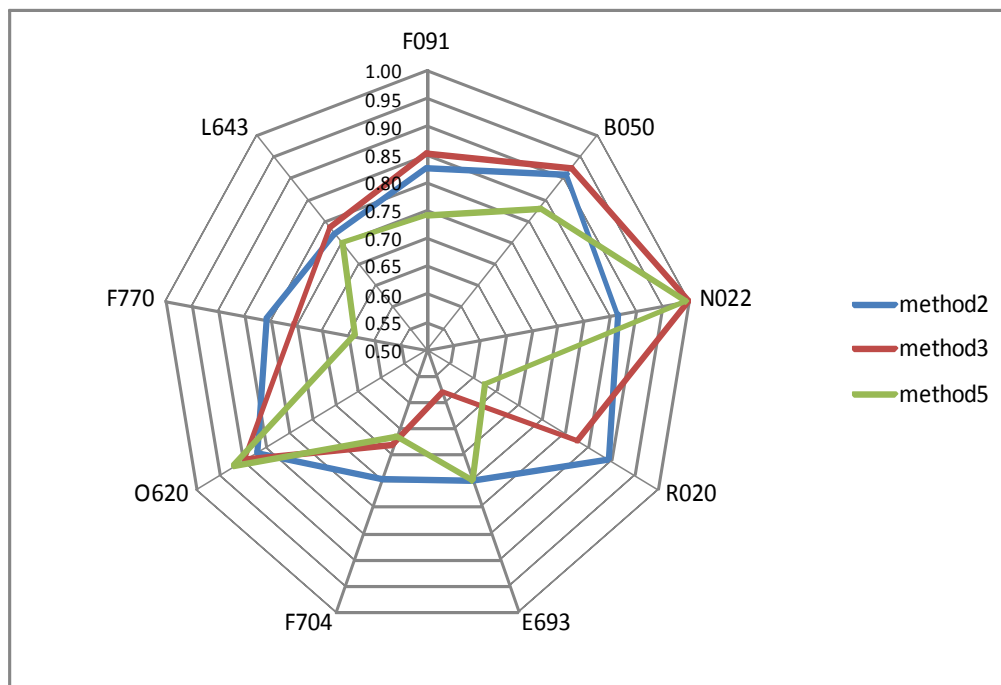


그림 20. 열외군 비율이 낮은 극단치 처리 방법들 간의 일치도(ROC).

나. KDRG 극단치 처리 방법의 적정성 검토 결과

건강보험 진료비 자료를 분석한 결과 분포가 확연하게 정규분포를 따르지 않는다는 것을 확인할 수 있었다. 질병군 진료비 분포의 정규 적합도를 확인하기 위해 Kolmogrov-Smirnov test를 시행한 결과 유의확률값이 유의수준 0.05값보다 작은 값으로 나타나 정규성을 따르지 않았다. 히스토그램을 통해 분포를 확인하였을 때도 positively skewed 형태를 보였으며, probability plot에서도 진료비를 나타내는 점들이 직선으로부터 이탈되었고, 극단값 또한 존재함을 확인할 수 있었다. 즉, 기본 질병군 단계인 ADRG 수준에서 건강보험 청구 자료의 분포가 양의 왜도를 가진 치우친 분포라는 특성을 파악할 수 있었다.

환자분류체계에서 자원소모의 검토 대상으로 평균 진료비와 함께 평균 재원일수를 검토하게 된다. 극단치의 일치정도를 파악하기 위한 분석변수 대상인 진료비와 재원일수의 실효성을 확인하기 위해 산점도와 상관분석을 실시하였다. 질병군별 산점도에서 진료비와 재원일수는 서로 양의 상관관계 형태를 보였으며, 피어슨 상관분석 결과 상관계수는 질병군별로 0.57~0.93으로 양의 상관관계를 나타냈으며, 유의확률값 또한 유의수준 0.05보다 작은값으로 나타나 두 변수간 양의 상관관계가 있다고 나타났다.

질병군의 빈도와 분포를 고려하여 9개 질병군을 대상으로 진료비와 재원일수 자료를 이용하여 극단값 판단에 대한 분할표를 생성하여 열외군 비율과 일치도를 측정하였다. method1과 method4를 질병군 별로 적용하였을 때 열외군 비율이 상대적으로 높게 나타나 두가지 방법은 제외하였다. 다음 단계로 method2, method3, method5를 대상으로 하여 일치정도를 ROC curve 지수로 상대 비교한 결과 method2와 method3 방법이 가장 일치도가 높게 나타났다. 그 중에서도 method2가 method3에 비해 질병군별 일치도 측면에서 변동의 폭이 작고 안정적이게 나타났다.

이상의 연구결과를 통해 method2 방법이 기타방법에 비해 진료비와 재원일수를 이용하여 극단치를 제외하였을 때 대부분의 질병군에서 극단치 비율이 5%미만으로 안정적이었으며, 일치도도 9개의 질병군 중 4개의 질병군에서 가장 높게 나타났다.

V. 고 찰

본 연구는 한국형 입원환자분류체계(KDRG)의 극단치 처리방법 적정성을 검토하기 위해 건강보험심사평가원에 청구된 2014년 1월 1일부터 2014년 12월 31일까지의 건강보험 입원환자 전체 진료건을 대상으로 시행하였다. 진료비 청구자료는 연간 진료내역 정보를 가진 대규모 전산자료로 KDRG의 극단치 처리 방법 적정성을 측정함에 있어 대표성을 가진 자료라고 볼 수 있다.

본 연구 진행과정에서 다음과 같이 몇 가지의 제한점 및 고찰이 있었다.

첫째, 극단치 처리 방법에서 현재의 연구는 질병군 별 진료비와 재원일수에 극단치 처리 방법들을 적용하여 열외군과 비열외군에 대해 상대적으로 가장 큰 일치를 나타내는 방법이 가장 만족스러운 극단치 처리 방법이라는 가설을 기반으로 한 제한점이 있다.

둘째, 열외군 비율 설정으로 인한 한계점이 있다. ADRG별 적정 수준 열외군 비율을 유지하기 위해서는 ADRG별 각 분포에 맞는 극단치 처리 방법을 적용하는 것이 타당하다. 획일적인 방법을 적용할 때 저빈도 질병군에서 열외군 비율이 10%를 초과하는 경우가 많거나, 분포가 모두 다른 ADRG별에서 열외군 비율이 0~10%사이값에서 그 이상 값까지 다양한 양상을 보이는 것은 당연하다. 자료가 큰 경우에도 10%이상 열외군 존재는 당연할 수 있으며 극단치 제거 목적(noise, cluster 개념)에 따라 각기 다른 방안을 모색해 볼 수 있다.

셋째, 방법론 설정의 한계점이 있다. 본 연구에서는 현행 극단치 처리 방법과 현행 방법에서 파생된 방법 및 일반적으로 많이 사용되는 극단치 처리 방법들을 대상으로 하였으며 극단값 제거 목적으로 상수값을 3으로 고정하였다. 현 연구에 사용된 방법 이외에도 다양한 방법론 설정과 함께 상수값을 3이외 1.5~3.0까지 다양하게 적용해 볼 수 있다.

넷째, 한국형 외래환자분류체계(KOPG)의 극단치 제거 방법의 경우 여러 극단치 처리 방법론을 대상으로 질병군별 데이터 손실율과 분산의 상대적 감소 정도를 나타내는 RIV(Reduction in Variance)를 확인하여 KDRG의 현행 극단치 처리 방법 2단계만을 적용하고 있다. 극단치로 제외된 자료의 비율이 높을수록 질병군의 동질성이 높아지게 되고 RIV값 또한 높아지게 된다. 따라서 RIV값이 극단치 제거 방법 선택에서 판단의 기준으로 활용되기에는 한계가 있을 수 있으며, KDRG 2단계 방법인 기하평균과 표준편차를 활용한 부분의 표기 및 수식에서도 추가적 연구가 필요하다.

다섯째, 동 기준 설정은 환자분류체계 개발 및 개정에 초점을 맞춘 제한점이 있다. 지불체계에서 총재정의 규모를 고려한 열외군 설정과 동일한 것은 아니다. 지불체계 내 열외군 설정 또는 추가보상(additional payment)은 별도의 고려 영역이다. 현재 자료원에는 원가(Cost)자료가 없으므로 자원소모변수를 총진료비(Charge)와 재원일수(LOS)로 하여 분석하였다. 유럽 국가의 규제 기관들은 효율적인 원가계산시스템 없이는 병원 보상 책정 방식으로 DRG를 제대로 적용 할 수 없음을 인식하였다. 즉 효과적이고 공정한 DRG 지불제도는 정확하고 수준 높은 원가계산시스템에 의해 크게 좌우된다.

본 연구는 환자분류체계에서 동질한 그룹 구성을 위해 극단치를 처리하는 목적으로 실시한 연구이지만 추후 지불체계에서 극단치를 고려 할 경우 투명하고 수준높은 병원의 원가 정보가 필요할 것이다. 앞으로 KDRG version이 개정되면서 질병군의 변화와 함께 지불체계를 고려한 환자분류체계의 개발에 관한 부분도 더 많은 연구가 필요할 것으로 생각된다.

VI. 결 론

본 연구는 2014년 건강보험 입원 청구 자료를 이용하여 현재 사용되고 있는 한국형 입원환자분류체계(KDRG)의 극단치 처리 방법의 문제점을 파악하고, 진료비 자료의 특성에 더욱 적합한 극단치 처리 방법을 모색하기 위한 목적으로 수행되었다. 건강보험 입원 청구 자료의 특성상 진료비 분포는 정규분포를 따르지 않고 오른쪽으로 꼬리가 늘어져있는 형태이며 대부분의 질병군에서 극단치가 존재한다. 현재 환자분류체계에서 극단치 제거시 2단계 누적형식으로 적용하고 있어 많은 질병군에서 열외군 비율이 높아져 데이터 손실의 문제점이 있었다.

연구 자료는 2014년 1월 1일부터 12월 31일까지 EDI로 청구된 행위별 건강보험 입원진료 전체 건으로, 총 건수는 7,691,707건이고 분석은 질병군 분류단계인 ADRG 수준에서 빈도와 분포를 고려하여 9개의 질병군을 대상으로 실시하여 비교하였다. 본 연구의 주요 결과는 다음과 같다.

첫째, 건강보험 진료비 자료의 특성 분석 및 분포 탐색 결과 선발된 9개의 ADRG 질병군에서 진료비 분포가 정규분포를 따르지 않았으며, 9개 질병군에 극단치 처리 방법들을 적용할 시 모두 극단값이 존재하는 것으로 확인되었다.

둘째, 환자분류체계에서 자원소모 검토시 평균진료비를 산출하고 추가적으로 평균재원일수도 함께 산출한다. 재원일수는 신 의료기술의 발전으로 인해 자원소모량의 검토 변수로 점점 의미가 사라지고 있으나, 급여자료 위주로 구축된 진료비 자료의 한계를 보완하기 위해 함께 사용된다. 분석변수 대상의 실효성을 확인하기 위해 질병군별 진료비와 재원일수의 상관분석을 실시한 결과 9개 질병군 모두 진료비와 재원일수간 양의 상관관계가 있음을 알 수 있었다.

셋째, 빈도와 분포를 고려하여 선정된 9개의 질병군을 대상으로 5가지 극단치 처리 방법을 적용하여 열외군 비율 정도와 진료비 및 재원일수 극단치의 일치도를 파악한 결과 현행 방법은 다른 방법에 비해 상대적으로 열외군 비율도 높았으며, 일치도 또한 낮은 것을 알 수 있었다.

이와 같이 자료의 빈도와 분포 등 질병군 자료의 특성 파악이 중요함을 알 수 있었으며, 앞으로 우리나라의 분류체계 극단치 처리 방법 적용시 극단치들을 많이 걸러내기 위해 2단계 누적방식의 현행방법을 적용하기 보다는, 빈도별 분포를 고려하고 데이터 손실정도를 파악하는 단계가 선행되어야 할 것이다. 분포의 특성에 맞는 극단치 처리방법의 선정 시 두가지 요소 즉, 데이터 손실율은 최소화하면서 진료비와 재원일수의 극단치 일치도를 높이는 방안을 함께 고려해 볼 수도 있다. 결론적으로 극단치 제거 방법 선정시 질병군별 빈도를 고려하고, 분포에 맞는 극단치 처리 방법을 적용하는 것이 타당하다. 또한 KDRG 활용 목적 및 극단치 처리 목적에 따라 각기 다른 방안을 모색하는 것도 필요하다.

본 연구는 향후 KDRG 차기 version 개정의 극단치 처리 방법으로 참고할 수 있을 것이다. KDRG 극단치 처리에 관한 선행 연구가 국내에 없는 상황에서 본 연구가 수행되며 많은 제한점이 있었다. 앞으로 KDRG에 관한 더 많은 연구가 필요할 것으로 생각된다.

참 고 문 헌

- 가와부치 코우이치. DRG/PPS와 병원경영. 서울: 한국의료컨설팅, 2000.
- 강길원. 한국형 진단명기준환자군 분류체계의 개선과 평가 - 소화기계 질병군을 중심으로[박사학위 논문]. 서울: 서울대학교 대학원; 2003.
- 건강보험심사평가원. 독일 환자분류체계 관련 국외출장 결과보고. 2012.
- 건강보험심사평가원. 환자분류체계 국외연수. 2013.
- 건강보험심사평가원. 국외출장 전달교육-호주연수 및 영국출장. 2014.
- 건강보험심사평가원. 한의 입원환자분류체계 평가연구. 2014.
- 건강보험심사평가원. KDRG version 3.5 분류집. 2014.
- 건강보험심사평가원. 2015년 국외출장 전달교육. 2015.
- 건강보험심사평가원. 환자분류체계 통계전문가 자문회의 결과 보고. 2015.
- 건강보험심사평가원. KDRG 중증도 개선을 위한 세미나 자료(AR-DRG V8.0개정 관련). 2015.
- 건강보험심사평가원. 업무매뉴얼 한 번 보면 이해가는 KDRG. 2016.

건강보험심사평가원. KDRG version 4.0 분류집. 2016.

건강보험심사평가원. KDRG version 4.1 분류집. 2017.

김형선. 한국형진단명기준환자군(KDRG) 분류체계의 동질성 평가[박사학위 논문], 서울: 이화여자대학교 대학원; 2012.

대한의사협회. 한국형 입원환자분류체계 V4.0 개발을 위한 임상적 타당성 연구. 2013.

박하영, 강길원, 고영. 한국형 외래환자분류체계의 개발과 평가. 보건행정학회지 2006;16(1):17-40.

신영수, 이영성, 박하영, 염용권. 한국형진단명기준환자군의 개발과 평가 : 입원환자의 의료서비스 이용을 중심으로. 예방의학회지 1993;26(2):293-309.

한국보건산업진흥원. 건강보험 DRG 지불제도 개선연구. 보건복지부, 2001.

한국보건산업진흥원. 1단계 DRG 개발 조정안. 보건복지부, 2001.

한국보건산업진흥원. 의료급여 DRG지불제도 연구(1). 보건복지부, 2001.

한국보건산업진흥원. 의료급여 DRG지불제도 도입 연구(2). 보건복지부, 2002.

- Busse R, Geissler A, Quentin W, Wiley M. Diagnosis-Related Groups in Europe. New York: Open University Press, 2011.
- Cots F, Elvira D, Castells X, Saez M. Relevance of outlier cases in case mix systems and evaluation of trimming methods. *Health Care Manag Sci.* 2003 Feb;6(1):27-35.
- Fantini MP, Cisbani L, Manzoli L, Vertrees J, Lorenzoni L. On the use of administrative databases to support planning activities: the case of the evaluation of neonatal case-mix in the Emilia-Romagna region using DRG and APR-DRG classification systems. *Eur J Public Health.* 2003 Jun;13(2):138-45.
- Freitas A, Silva-chargea T, Lopes F, Garcia-Lema I, Teixeira-Pinto A, Brazdil P, chargea-Pereira A. Factors influencing hospital high length of stay outliers. *BMC Health Serv Res.* 2012 Aug 20;12:265.
- Gaal P, Stefka N, Nagy J. charge accounting methodologies in price setting of acute inpatient services in Hungary. *Health Care Manag Sci.* 2006 Aug;9(3):243-50.
- Marazzi A, Paccaud F, Ruffieux C, Beguin C. Fitting the distributions of length of stay by parametric models. *Med Care.* 1998 Jun; 36(6): 915-27.
- Palmer G, Reid B. Evaluation of the performance of diagnosis-related groups and similar casemix systems: methodological issues. *Health Serv Manage Res.* 2001 May;14(2):71-81.

Pirson M, Dramaix M, Leclercq P, Jackson T. Analysis of charge outliers within APR-DRGs in a Belgian general hospital: two complementary approaches. *Health Policy*. 2006 Mar;76(1):13-25.

Reid B, Palmer G, Aisbett C. The performance of Australian DRGs. *Aust Health Rev*. 2000; 23(2):20-31.

Reid B, Sutch S. Comparing diagnosis-related group systems to identify design improvements. *Health Policy*. 2008 Jul;87(1):82-91.

SAS global community

(URL: <http://sascommunity.org/sugi/SUGI89/Sugi-89-265%20Bretheim.pdf>).

Schreyogg J, Stargardt T, Tiemann O, Busse R. Methods to determine reimbursement rates for diagnosis related groups (DRG): a comparison of nine European countries. *Health Care Manag Sci*. 2006 Aug;9(3): 215-23.

Thomas JW, Ward K. Economic profiling of physician specialists: use of outlier treatment and episode attribution rules. *Inquiry*. 2006 Fall; 43(3):271-82.

Wikimedia commons(URL: <http://commons.wikimedia.org>).

= ABSTRACT =

**A study of outlier trimming methods for KDRG
(Korean Diagnosis-Related Group)**

Jinhwi Kim

Department of Biostatistics

Graduate School of Public Health

Yonsei University

(Directed by Professor Chungmo Nam, Ph.D.)

Background

KDRG is a type of PCS(Patient Classification System) that classifies acute inpatients into the group with clinical similarity and the homogeneity of resources, using information, such as diagnosis and procedures of patients. Current standards on detection and treatment of KDRG outliers are developed based on methods of RDRG(Refined DRG) in U.S. at the time of developing KDRG version 2.0(1991) and the same methods have been applied to KDRG so far without revising. As KDRG has continually revised, it has become more important and been needed to make new standards of dealing with outliers. Therefore, this

study describes the problems of methods currently used for KDRG and seeks for outlier treatment methods that are more appropriate to characteristics of medical charge and LOS data in Korea.

Subject and Methods

The study is based on the fee-for-service claims of 7,691,707 database(KDRG version 4.0) that were submitted to the HIRA(Health Insurance Review and Assessment service) through EDI(Electronic Data Interchange) from January 2014 to December 2014. The subjects are 9 DRGs considering the frequency and distribution at the ADRG(Adjacent DRG) level. After exploring relationship and distribution of the medical charge and the LOS, various outlier treatment methods were applied and compared with the current method. Analysis was conducted by comparing statistics about the rate of trimmed outliers, and also the correspondence between the charge outliers and LOS outliers.

Results

In order to research on the characteristics of the medical charge and the LOS data, the distribution of data was identified through descriptive statistics and goodness fit test for a normal distribution. As the result, the distribution of the medical charge and the LOS did not follow the normal distribution and were right-skewed. Correlation analysis was performed to identify effective variables and there was a strong positive

correlation between the medical charge and the LOS from 0.57 to 0.93. As a result of applying 5 outlier treatment methods including the current one to 9 DRGs, the rate of trimmed outliers in most of DRGs was exceed 5% when the current one was applied. Also, the correspondence between the charge outliers and the LOS outliers was relatively lower than the other methods. However, when only the 1-step of current method using quartiles and interquartile range was applied, it was found that the proportion of trimmed outliers was relatively low and the correspondence was high.

Conclusions

With regard to appropriate outlier treatment methods for KDRG, considering the distribution by frequency and the amount of data loss is priorly needed before trimming outliers. Among outlier treatment methods fit the distribution, the data loss rate and the correspondence between the charge outliers and the LOS outliers might be considered, since the LOS is also an important variable. When choosing outlier treatment methods for KDRG, considering the frequency of DRGs and applying outlier treatment methods fit the distribution should be appropriate. Moreover, the most important thing is to seek proper methods according to the use of KDRG and the purpose of outlier treatment.

Key words : DRG, KDRG, outlier, trim